

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:08:53 ; Search time 16 Seconds
(without alignments)
54.108 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	347	2 S10571	mucin 1 precursor,
2	49	100.0	1344	1 A35175	mucin 1 precursor,
3	45	91.8	256	2 A60533	tumor-associated a
4	41	83.7	590	2 AF2355	hypothetical prote
5	38	77.6	717	2 S38177	SSV7 protein homol
6	37	75.5	1183	2 T39233	probable inositol
7	36	73.5	329	2 T13016	hypothetical prote
8	36	73.5	335	2 A70851	probable trxB2 pro
9	35	73.5	690	2 S54775	cell size regulati
10	35	71.4	154	2 S39873	hypothetical prote
11	35	71.4	182	2 T20533	hypothetical prote
12	35	71.4	347	2 T29742	hypothetical prote
13	35	71.4	363	2 T46907	hypothetical prote
14	35	71.4	653	2 E84682	hypothetical prote
15	35	71.4	760	1 S07896	transcription fact
16	34	69.4	145	2 H69051	heterodisulfide re
17	34	69.4	190	2 T04543	hypothetical prote
18	34	69.4	275	2 T12937	probable phosphoe
19	34	69.4	312	2 A86279	F14L17.21 protein
20	34	69.4	340	2 T40392	probable 3-beta-hy
21	34	69.4	604	2 T37870	RNA-binding / Ran
22	34	69.4	1006	2 E87363	medium-chain-fatty
23	34	69.4	1792	2 A57075	tensin - chicken {
24	34	69.4	2481	2 A43908	fibronectin - Afri
25	33	67.3	135	2 C47114	phosphoprotein pho
26	33	67.3	174	2 F85028	hypothetical prote
27	33	67.3	267	2 S21399	hypothetical prote
28	33	67.3	284	2 F51172	transcription fact
29	33	67.3	291	2 A98325	dipeptide ABC tran

30 33 67.3 291 2 AE2958
31 33 67.3 481 2 T39357
32 33 67.3 551 2 T00984
33 33 67.3 727 2 C84534
34 33 67.3 850 2 AH0359
35 33 67.3 1844 1 RRWETM
36 33 67.3 2140 2 T18543
37 32 65.3 92 2 AC0406
38 32 65.3 94 2 AB0949
39 32 65.3 122 2 T46369
40 32 65.3 206 2 T23852
41 32 65.3 211 2 B23162
42 32 65.3 215 2 T18899
43 32 65.3 235 2 JQ2260
44 32 65.3 240 2 H89692
45 32 65.3 249 2 S21935

ALIGNMENTS

RESULT 1

S10571
mucin 1 precursor, secreted epithelial tumor antigen splice form - human
N;Contains: mucin 1 secreted breast-cancer-associated splice form
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 01-Dec-2000
C;Accession: S10571; JN0100; I56024; S09706; S10217
R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.;
Eur. J. Biochem. 189, 463-473, 1990
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may generat
A;Reference number: S10571; MUID:90276413; PMID:2351132
A;Accession: S10571
A;Molecule type: mRNA
A;Residues: 1-347 <WRE>
A;Cross-references: EMBL:X52228; NID:G36434; PIDN:CAA36477.1; PID:G36435
R;Tsarfaty, I.; Hareuveni, M.; Horev, J.; Zaretsky, J.; Weiss, M.; Jeltsch, J.M.; Garnie
Gene 93, 313-318, 1990
A;Title: Isolation and characterization of an expressed hypervariable gene coding for a l
A;Reference number: JN0100; MUID:91033045; PMID:1698329
A;Accession: JN0100
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-135, 'Q', 137-142, 'E', 144-163, 204-208, 'A', 210-347 <TSA>
A;Cross-references: GB:M35093; NID:G182252; PIDN:AA859612.1; PID:G182253
R;Xing, P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie, J
J. Immunol. 142, 3503-3509, 1989
A;Title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.
A;Reference number: I56024; MUID:89235154; PMID:2715633
A;Accession: I56024
A;Status: preliminary; translated from GB/EMBL/DDRJ
A;Molecule type: mRNA
A;Residues: 182-201 <RES>
A;Cross-references: GB:M26316; NID:G516622; PIDN:AAA36336.1; PID:G516623
R;Tendler, S.J.B.
Biochem. J. 267, 733-737, 1990
A;Title: Elements of secondary structure in a human epithelial mucin core peptide fragme
A;Reference number: S09706; MUID:90253387; PMID:2339983
A;Accession: S09706
A;Molecule type: protein
A;Residues: 182-201 <TEN>
C;Genetics:
A;Gene: GDB:MUC1; PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map position: lq21-lq23
C;Keywords: alternative splicing; tandem repeat
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-347/Product: mucin 1, secreted epithelial tumor antigen splice form #status predict
F;24-163,204-347/Product: mucin 1, secreted breast-cancer-associated splice form #status

Query Match 100.0%; Score 49; DB 2; Length 347;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
 |||||
 Db 210 STAPPVHV 218

RESULT 2

mucin 1 precursor, repetitive splice form A [validated] - human
 N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin
 noreatic mucin; polymorphic epithelial mucin [PEM]
 N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,
 C:Species: Homo sapiens (man)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
 A:Accession: A35175; B35175; A35886; A35887; S10572; A36735; PX0066; S10218; S51
 R:Biochem. M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
 J. Biol. Chem. 265, 5573-5578, 1990
 A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
 A:Reference number: A35175; MUID:90202794; PMID:2318825
 A:Accession: A35175
 A:Molecule type: mRNA
 A:Residues: 1-952,1033-1344 <LIG1>
 A:Cross-references: GB:M32738; GB:J05288; NID:G182121; PIDN:AAA35804.1; PID:G182124; GB:
 A:Experimental source: splice form A
 A:Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl-er
 A:Accession: B35175
 A:Molecule type: mRNA
 A:Residues: 1-19,29-952,1033-1344 <LIG2>
 A:Cross-references: GB:M32739; GB:J05288; NID:G182126; PIDN:AAA35806.1; PID:G182129; GB:
 A:Experimental source: splice form B
 A:Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-er
 R:Gendler, S.J.; Lancia, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchell
 J. Biol. Chem. 265, 15286-15293, 1990
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli
 A:Reference number: A35886; MUID:90368715; PMID:1697589
 A:Accession: A35886
 A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05581; NID:G188869; PIDN:AAA59876.1; PID:G188870
 A:Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
 R:ian, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
 J. Biol. Chem. 265, 15294-15299, 1990
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
 A:Reference number: A35887; MUID:90368716; PMID:2394722
 A:Accession: A35887
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:G189599
 A:Note: GenBank entry HUMPANMU containing four fewer copies of the tandemly repeated sequ
 R:Wreschner, D.H.; Hareuveni, M.; Tearfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
 Eur. J. Biochem. 189, 463-473, 1990
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
 A:Reference number: S10571; MUID:90276413; PMID:2351132
 A:Accession: S10572
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WRE>
 A:Cross-references: EMBL:X52229; NID:G37053
 R:Wreschner, D.H.
 submitted to the EMBL Data Library, March 1990
 A:Reference number: S40293
 A:Accession: S40293
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <WR2
 A:Cross-references: EMBL:X52229; NID:G37053; PIDN:CAA36478.1; PID:G37054
 R:Abé, M.; Siddiqui, J.; Kufe, D.
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473; PMID:2597151
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>

A:Cross-references: EMBL:M31823; NID:G181542; PIDN:AAA35757.1; PID:G181543
 R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
 A:Reference number: JX0235; MUID:93123189; PMID:1478919
 A:Accession: PX0066
 A:Molecule type: mRNA
 A:Residues: 998-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
 A:Experimental source: gastric carcinoma cell
 R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 356, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
 A:Reference number: S51026; MUID:95080414; PMID:7988707
 A:Contents: annotation
 A:Note: undertermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
 partial repeats. The repeat shown is defined by SmaI nuclease sites.
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
 C:Genetics:
 C:Gene: GDB:MUC1; PDM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q21-1q23
 A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 C:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
 F1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRA>
 F1-62/Region: mucin 1 amino-terminal non-repetitive
 F1-23/Domain: signal sequence #link PRA #status predicted <SIGA>
 F1-19,29-32/Domain: signal sequence #link PRA #status predicted <SIGB>
 F1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
 F1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
 F138-1017/Region: 20-residue repeats (GSTAPPAGVTSADTRAP)
 F143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F1245-1272/Domain: transmembrane #status predicted <TRW>
 F1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F1213/Binding site: phosphate (Tyr) (covalent) #status predicted
 Query Match 100.0%; Score 49; DB 1; Length 1344;
 Best Local Similarity 100.0%; Pred. No. 0.53;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHV 9
 |||||
 Db 1039 STAPPVHV 1047
 RESULT 3
 A60533
 tumor-associated antigen DF3 - human
 C:Species: Homo sapiens (man)
 C:Date: 19-Mar-1993 #sequence_revision 07-May-1993 #text_change 24-Nov-1999
 C:Accession: A60533
 R:Merlo, G.R.; Siddiqui, J.; Cropp, C.S.; Liscia, D.S.; Lidereau, R.; Callahan, R.; Kufe
 Cancer Res. 49, 6966-6971, 1989
 A:Title: Frequent alteration of the DF3 tumor-associated antigen gene in primary human b
 A:Reference number: A60533; MUID:90058554; PMID:2582438
 A:Accession: A60533
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-256 <MER>
 C:Genetics:
 C:Map position: 1q21-q24
 C:Superfamily: proline-rich protein
 C:Keywords: glycoprotein; tandem repeat
 Query Match 91.8%; Score 45; DB 2; Length 256;
 Best Local Similarity 100.0%; Pred. No. 0.52;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHV 8
 |||||
 Db 193 STAPPVHV 200


```

RESULT 4
AF2555
hypothetical protein all8037 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120ga
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2555
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <KUR>
A:Cross-references: GB:AF003603; PIDN:BAB77367.1; PID:gl7134810; GSPDB:GN00182
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all8037
A:Genome: plasmid

Query Match      83.7%; Score 41; DB 2; Length 590;
Best Local Similarity 87.5%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TAPPVHNV 9
Db      9 TAPPAHNV 16

RESULT 5
S38177
SSV7 protein homolog - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKR098C
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 06-Feb-1998
C:Accession: S38177
R:Gaillon, L.; Dujon, B.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38175
A:Accession: S38177
A:Molecule type: DNA
A:Residues: 1-717 <GAI>
A:Cross-references: EMBL:Z28323; NID:5486592; PID:5486593; MIPS:YKR098C
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:URP11
A:Cross-references: SGD:S0001806; MIPS:YKR098C
A:Map position: 11R

Query Match      77.6%; Score 38; DB 2; Length 717;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      1 STAPPVHNV 9
Db      231 ATAPPVHSL 239

RESULT 6
T39233
probable Inositol polyphosphate phosphatase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39233
R:Churcher, C.M.; Gentles, S.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: 221837
A:Accession: T39233
A:Status: preliminary; translated from GB/EMBL/DBJ

```

```

A:Molecule type: DNA
A:Residues: 1-1183 <CHU>
A:Cross-references: EMBL:Z98763; PIDN:CAB11494.1; GSPDB:GN00066; SPDB:SPAC9G1.10c
A:Experimental source: strain 972h; cosmid c9G1
C:Genetics:
A:Gene: SPDB:SPAC9G1.10c
A:Map position: 1

Query Match      75.5%; Score 37; DB 2; Length 1183;
Best Local Similarity 77.8%; Pred. No. 83;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 STAPPVHNV 9
Db      253 STAPPVHNI 261

RESULT 7
TI3016
hypothetical protein F8L21.40 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 22-Oct-1999
C:Accession: TI3016
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, July 1999
A:Reference number: Z17587
A:Accession: TI3016
A:Molecule type: DNA
A:Residues: 1-329 <BEV>
A:Cross-references: EMBL:AL096882; GSPDB:GN00062; ATSP:F8L21.40
A:Experimental source: cultivar Columbia; BAC clone F8L21
C:Genetics:
A:Gene: ATSP:F8L21.40
A:Map position: 4

Query Match      73.5%; Score 36; DB 2; Length 329;
Best Local Similarity 55.6%; Pred. No. 33;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      1 STAPPVHNV 9
Db      307 SSJPLHNI 315

RESULT 8
A70851
probable trxB2 protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70851
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome :
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70851
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-335 <COL>
A:Cross-references: GB:AL021426; GB:AL123456; NID:g3261511; PIDN:CAA16226.1; PID:g280869;
C:Superfamily: thioredoxin reductase; thioredoxin reductase homology
F:11-315/Domain: thioredoxin reductase homology <TRXB>
C:Genetics:
A:Gene: trxB2

Query Match      73.5%; Score 36; DB 2; Length 335;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 TAPPVHNV 8

```



```

db      |||||:
        2 TAPPVHD 8

RESULT 9
S54775
cell size regulation protein RCS1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: AFT1 protein; protein G3266; protein YGL071w
C:Species: Saccharomyces cerevisiae
C:Date: 27-Oct-1995 #sequence revision 03-Nov-1995 #text_change 29-Oct-1999
C:Accession: S54775; S54874; S12915; S64078
R:Yamaguchi-Iwai, Y.; Dancis, A.; Klausner, R.D.
EMBO J. 14, 1231-1239, 1995
A:Title: AFT1: a mediator of iron regulated transcriptional control in Saccharomyces cerevisiae
A:Reference number: S54775; MUID:95237204; PMID:720713
A:Accession: S54775
A:Molecule type: mRNA
A:Residues: 1-690 <YAM>
A:Cross-references: EMBL:Z48004; NID:g758277; PIDN:CAA88044.1; PID:g758278
R:Aldea, M.; Casas, C.; Gallego, C.; Espinet, C.; Herrero, E.
submitted to the EMBL Data Library, January 1994
A:Reference number: S54874
A:Accession: S54874
A:Molecule type: DNA
A:Residues: 1-7, 'H', 9-70, 'S', 72-135, 'T', 137-235, 'N', 237-415, 'G', 417-506, 'S', 508-537, 'P', 569-90 <AUD>
A:Cross-references: EMBL:X77413; NID:g840880; PIDN:CAA54586.1; PID:g840881
R:Gil, R.; Zueco, J.; Santandreu, R.; Herrero, E.
Yeast 7, 1-14, 1991
A:Title: RCS1, a gene involved in controlling cell size in Saccharomyces cerevisiae.
A:Reference number: S12915; MUID:91213643; PMID:2021081
A:Accession: S12915
A:Molecule type: DNA
A:Residues: 370-377, 'L', 379-415, 'G', 417-467, 'T', 469-506, 'S', 508-537, 'L', 539-567, 'P', 569-90 <RIE>
A:Cross-references: EMBL:X53046; NID:g4294; PIDN:CAA37215.1; PID:g4295
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64071
A:Accession: S64071
A:Molecule type: DNA
A:Residues: 1-690 <RIE>
A:Cross-references: EMBL:Z72593; NID:g1322582; PIDN:CAA96775.1; PID:e243959; PID:g132258
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:RCS1; AFT1
A:Cross-references: SGD:S0003039; MIPS:YGL071w
A:Map position: 7L
C:Function:
A:Description: control of iron uptake

Query Match      73.5%; Score 36; DB 2; Length 690;
Best Local Similarity 75.0%; Pred. No. 72;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 STAPPVHN 8
        |||||:
        497 SSAPPVHH 504

Db

RESULT 10
S39873
hypothetical protein 154 - Streptomyces ambofaciens
C:Species: Streptomyces ambofaciens
C:Date: 19-May-1994 #sequence_revision 26-May-1995 #text_change 26-May-2000
C:Accession: S39873; S33421
R:Hagege, J.; Pernodet, J.L.; Friedmann, A.; Guerin, M.
Mol. Microbiol. 10, 799-812, 1993
A:Title: Mode and origin of replication of pSAM2, a conjugative integrating element of S. ambofaciens
A:Reference number: S39873; MUID:95020551; PMID:7934842
A:Accession: S39873
A:Molecule type: DNA
A:Residues: 1-154 <HAG>
A:Cross-references: EMBL:Z19590; NID:g298044; PIDN:CAA79638.1; PID:g298045

db      |||||:
        2 TAPPVHD 8

Query Match      71.4%; Score 35; DB 2; Length 154;
Best Local Similarity 55.6%; Pred. No. 23;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy      1 STAPPVHN 9
        |||||:
        11 ATSPPLHSV 19

Db

RESULT 11
T20533
hypothetical protein F07A5.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T20533
R:Wilkinson, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19288
A:Accession: T20533
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-182 <WIL>
A:Cross-references: EMBL:Z72506; PIDN:CAA96615.1; GSPDB:GN000019; CBSP:F07A5.5
A:Experimental source: clone F07A5
C:Genetics:
A:Gene: CBSP:F07A5.5
A:Map position: 1
A:Introns: 19/3; 57/1; 80/1; 115/2; 153/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F07A5.5

Query Match      71.4%; Score 35; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 TAPPVH 7
        |||||:
        163 TAPPVH 168

Db

RESULT 12
T29742
hypothetical protein ZC317.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29742
R:Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid ZC317.
A:Reference number: Z20677
A:Accession: T29742
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-347 <GEI>
A:Cross-references: EMBL:U64840; PIDN:AAE04959.1; GSPDB:GN000023; CBSP:ZC317.3
A:Experimental source: strain Bristol N2; clone ZC317
C:Genetics:
A:Gene: CBSP:ZC317.3
A:Map position: 5
A:Introns: 20/2; 69/2; 124/3; 228/3; 277/1; 310/3
C:Superfamily: acetylcholine receptor

Query Match      71.4%; Score 35; DB 2; Length 347;
Best Local Similarity 55.6%; Pred. No. 54;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      1 STAPPVHN 9
        |||||:
        324 STGFSIHN 332

Db

```

A>Note: the authors translated the codon TAC for residue 81 as Thr and TAC for residue 111 as Ile.
C:Superfamily: unassigned mult domain proteins; mult domain homology
F:43-77/Domain: mult domain homology <MULT>

RESULT 13

T46907 hypothetical protein DKFZp761G2023.1 - human

C;Species: Homo sapiens (man)
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 17-Mar-2000

C;Accession: T46907

R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, February 2000

A;Reference number: Z24137

A;Accession: T46907

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-363 <AAA>

A;Cross-references: EMBL:AL136570

A;Experimental source: adult amygdala; clone DKFZp761G2023

C;Genetics:

A;Note: DKFZp761G2023.1

Query Match 71.4%; Score 35; DB 2; Length 363;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TAPPVH 7

|||||

337 TAPPVH 342

RESULT 14

E84682

hypothetical protein At2g28240 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: E84682

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.V.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Unayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, J.

Nature 402, 761-766, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: E84682

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-653 <STO>

A;Cross-references: GB:AE002093; NID:94803957; PIDN:AD29829.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g28240

A;Map position: 2

Query Match 71.4%; Score 35; DB 2; Length 653;
 Best Local Similarity 55.6%; Pred. No. 1e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9

:|||||:

595 ASAPPVHGI 603

RESULT 15

S07896

transcription factor Oct-1.32 - African clawed frog

N;Alternate names: maternal transcription factor; pou homeobox protein oct-1.32

C;Species: Xenopus laevis (African clawed frog)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999

C;Accession: S07896; S40642; S08673; S14554; S12186

R;Smith, D.P.; Old, R.W.

Nucleic Acids Res. 18, 369, 1990

A;Title: Nucleotide sequence of Xenopus laevis Oct-1 cDNA.

A;Reference number: S07896; MUID:90221827; PMID:2326173

A;Accession: S07896

A;Molecule type: mRNA

A;Residues: 1-760 <SMI>

A;Cross-references: EMBL:X17190; NID:964942; PIDN:CAA35051.1; PID:9671665

R;Smith, D.P.; Old, R.W.
 Nucleic Acids Res. 19, 815-821, 1991
 A;Title: Xenopus laevis Oct-1 does not bind to certain histone H2B gene promoter octamer
 A;Reference number: S40642; MUID:91204435; PMID:2017364
 A;Accession: S40642
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-7, 'L', 9-760 <SM2>
 R;Schilthuis, J.G.; Baarends, W.M.; Peterson-Maduro, J.; Destre, O.H.J.
 submitted to the EMBL Data Library, February 1990

A;Reference number: S08673

A;Accession: S08673

A;Molecule type: DNA

A;Residues: 'LVCCSFLTIQYSV', 261-335 <SCH>

A;Cross-references: EMBL:X51819; NID:965249; PIDN:CAA36119.1; PID:9295726

R;Hinkley, C.; Leibham, D.; Perry, M.

submitted to the EMBL Data Library, January 1991

A;Description: Regulated binding specificity of Xenopus OCT-1, a maternal transcription

A;Reference number: S14554

A;Accession: S14554

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-156, S, 158-367, 370-760 <HIN>

A;Cross-references: EMBL:X57165; NID:964951; PIDN:CAA40454.1; PID:964952

R;Baltzinger, M.; Stiegler, P.; Remy, P.

Nucleic Acids Res. 18, 6131, 1990

A;Title: Cloning and sequencing of POU-boxes expressed in Xenopus laevis neurula embryos

A;Reference number: S12179; MUID:91045083; PMID:2235499

A;Accession: S12186

A;Molecule type: mRNA

A;Residues: 320-435 <BAL>

A;Cross-references: EMBL:X54683; NID:964947; PIDN:CAA38497.1; PID:9930281

C;Genetics:

A;Gene: Oct-1

A;Introns: 320/3

C;Superfamily: transcription factor Oct-1; homeobox homology; POU domain homology

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

F;301-368/Domain: POU domain homology <POU>

F;398-454/Domain: homeobox homology <HOX>

Query Match 71.4%; Score 35; DB 1; Length 760;

Best Local Similarity 77.8%; Pred. No. 1.2e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9

|||||:

529 STAPPVSV 537

Search completed: July 28, 2004, 18:11:08

Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:07:33 ; Search time 14 Seconds

(without alignments)
33.474 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	475	1 MUC1_HYLLA	Q9435 hylobates 1
2	49	100.0	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
3	38	77.6	1717	1 USP8_YEAST	P36026 saccharomyc
4	36	73.5	335	1 TXSB_MYCTU	P52214 mycobacteri
5	36	73.5	690	1 APT1_YEAST	P22149 saccharomyc
6	35	71.4	154	1 MUTT_STRAM	P32091 streptomyce
7	35	71.4	363	1 LH61_HUMAN	Q9upm6 homo sapien
8	35	71.4	363	1 LH61_MOUSE	Q9riro mus musculu
9	35	71.4	760	1 P021_XENLA	P16143 xenopus lae
10	34	69.4	2481	1 FINC_XENLA	Q91740 xenopus lae
11	33	67.3	1391	1 NCO3_XENLA	O57539 xenopus lae
12	33	67.3	1844	1 POLR_TYMYA	P20128 turnip yell
13	32	65.3	211	1 PR33_DAUCA	P06600 daucus caro
14	32	65.3	249	1 HL_DROHY	P17268 drosophila
15	32	65.3	265	1 COQ4_HUMAN	Q9y3a0 homo sapien
16	32	65.3	473	1 LCB1_CRIGR	O54695 cricetus
17	32	65.3	473	1 LCB1_MOUSE	Q35704 mus musculu
18	32	65.3	549	1 YST1_CABEL	Q22000 caenorhabdi
19	32	65.3	583	1 YP65_MYCTU	O50733 mycobacteri
20	32	65.3	584	1 PI43_MOUSE	O54714 mus musculu
21	32	65.3	633	1 IPAA_SHIFL	P18010 shigella fl
22	32	65.3	761	1 YKC2_CABEL	P41993 caenorhabdi
23	32	65.3	813	1 FOR2_XENLA	Q03364 xenopus lae
24	32	65.3	850	1 NRG2_HUMAN	O14511 homo sapien
25	32	65.3	1669	1 ASX_DROME	Q9y727 drosophila
26	32	65.3	1744	1 TENS_CHICK	Q94205 gallus gall
27	31	63.3	108	1 REC2_HUMAN	Q9y2y1 homo sapien
28	31	63.3	171	1 HMEB_XENLA	P31538 xenopus lae
29	31	63.3	174	1 CT27_HUMAN	Q9gzns homo sapien
30	31	63.3	193	1 YF91_MYCTU	O06599 mycobacteri
31	31	63.3	198	1 YF91_MOUSE	O49626 mycobacteri
32	31	63.3	228	1 RX1_CHICK	Q9pvy0 gallus gall
33	31	63.3	243	1 PSD_MYCLE	Q9ccw9 mycobacteri

34	31	63.3	261	1	KLK9_MOUSE	P15949 mus musculu
35	31	63.3	263	1	DRG1_RAT	Q62798 rattus norv
36	31	63.3	310	1	PR28_MYCTU	P71697 mycobacteri
37	31	63.3	330	1	RX1_ERARE	O42356 brachydaio
38	31	63.3	370	1	SERC_CABEL	P91856 caenorhabdi
39	31	63.3	441	1	HISX_STRCO	P16245 streptomyce
40	31	63.3	478	1	MOT2_HUMAN	O60669 homo sapien
41	31	63.3	551	1	PODX_RABIT	Q28645 oryctolagus
42	31	63.3	759	1	LIPS_MOUSE	P54310 mus musculu
43	31	63.3	768	1	LIPS_RAT	P15304 rattus norv
44	31	63.3	901	1	AL80_MOUSE	Q61548 mus musculu
45	31	63.3	907	1	AL80_HUMAN	O60641 homo sapien

ALIGNMENTS

RESULT 1
MUC1_HYLLA
ID MUC1_HYLLA STANDARD; PRT; 475 AA.
AC Q29435;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 1 precursor (MUC-1).
GN MUC1.
OS Hylobates lar (Common gibbon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX NCBI_TaxID=9580;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains";
RL Mamm. Genome 6:885-888(1995).
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC CYTOSKELETON (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -!- SIMILARITY: Contains 1 SEA domain.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L41589; AAA69965.1; -;
CC EMBL; L41625; AAA69918.1; -;
CC EMBL; L41624; AAA69918.1; JOINED.
CC InterPro; IPR000082; SEA_domain.
CC Pfam; PF01390; SEA; 1.
CC SMART; SMO0200; SEA; 1.
CC PROSITE; PS00024; SEA; 1.
CC Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
CC Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 475 MUCIN 1.
FT DOMAIN 24 380 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 381 401 POTENTIAL.
FT DOMAIN 402 475 CYTOPLASMIC (POTENTIAL).
FT REPEAT 102 121 1.
FT REPEAT 122 141 2.
FT REPEAT 142 161 3.
FT REPEAT 162 181 4.
FT DOMAIN 254 371 SEA.
SQ SEQUENCE 475 AA; 49371 MW; D7A699D6D68C6622 CRC64;

Query Match

100.0%; Score 49; DB 1; Length 475;

Best Local Similarity 100.0%; Pred. No. 0.091; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
|||||||
Db 170 STAPPVHV 178

RESULT 2

MUC1 HUMAN
ID MUC1 HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQL1; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=90202794; PubMed=2318825;
RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=91097524; PubMed=2269309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;

RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=9103045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Keydar I., Hilkens J., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen";
RL Gene 93:313-318(1990).
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RX MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]
RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
RX MEDLINE=97355747; PubMed=9212228;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/Z)";
RL Int. J. Cancer 72:87-94(1997).
RN [10]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RX Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A. (ISOFORM 9).
RX TISSUE=Epithelial cancer;
RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [13]
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [14]
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
RX TISSUE=Thyroid;
RA MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene";
RL Int. J. Cancer 66:55-59(1996).
RN [15]
RP SEQUENCE OF 1-89 FROM N.A.
RX TISSUE=Lung;
RA MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues";


```

RN  Oncology 53:118-126(1996).
RN  [16]
RP  SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
RC  TISSUE=Breast carcinoma.
RA  Bulwella L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RL  Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN  [17]
RP  CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX  MEDLINE=97460054; PubMed=9312074;
RA  Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
RA  Hanisch F.-G.;
RT  "Localization of O-glycosylation sites on glycopeptide fragments from
RT  lactation-associated MUC1. All putative sites within the tandem
RT  repeat are glycosylation targets in vivo."
RL  J. Biol. Chem. 272:24780-24793(1997).
RN  [18]
RP  CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX  MEDLINE=99303572; PubMed=10373415;
RA  Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
RA  Hanisch F.-G.;
RT  "High density O-glycosylation on tandem repeat peptide from secretory
RT  MUC1 of T47D breast cancer cells."
RL  J. Biol. Chem. 274:18165-18172(1999).
RN  [19]
RP  POLYMORPHISM WITHIN THE REPEAT.
RX  MEDLINE=21359366; PubMed=11350974;
RA  Engelmann K., Baldus S.E., Hanisch F.-G.;
RT  "Identification and topology of variant sequences within individual
RT  repeat domains of the human epithelial tumor mucin MUC1."
RL  J. Biol. Chem. 276:27764-27769(2001).
RN  [20]
RP  CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
RX  MEDLINE=99211485; PubMed=10197628;
RA  Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
RA  Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
RA  Wreschner D.H.;
RT  "The breast cancer-associated MUC1 gene generates both a receptor and
RT  its cognate binding protein."
RL  Cancer Res. 59:1552-1561(1999).
RN  [21]
RP  PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RX  MEDLINE=21240104; PubMed=11341784;
RA  Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
RA  Harris A.;
RT  "Identification of MUC1 proteolytic cleavage sites in vivo."
RL  Biochem. Biophys. Res. Commun. 283:715-720(2001).
RN  [22]
RP  CHARACTERIZATION
RX  MEDLINE=21836452; PubMed=11847293;
RA  Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
RA  Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
RA  Stacey M., Lin H.-H., Gordon S.;
RT  "Generation of ligand-receptor alliances by 'SEA' module-mediated
RT  cleavage of membrane-associated mucin proteins."
RL  Protein Sci. 11:698-706(2002).
RN  [23]
RP  PHOSPHORYLATION.
RX  MEDLINE=95080414; PubMed=7988707;
RA  Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
RT  "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
RT  Cytokine receptor-like molecules."
RL  FEBS Lett. 356:130-136(1994).
CC  -!- FUNCTION: May play a role in adhesive functions and in cell-cell
CC  interactions, metastasis and signaling. May provide a protective
Query Match 100.0%; Score 49; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 STAPPVHNV 9
Db 950 STAPPVHNV 958
[1]
RN  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1773;
RN  [1]
```

```

RESULT 3
UBPB YEAST STANDARD; PRT; 717 AA.
AC P36026;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 11 (BC 3.1.2.15) (Ubiquitin
DE thiolesterase 11) (Ubiquitin-specific processing protease 11)
DE (Deubiquitinating enzyme 11).
GN UBPI1 OR YKR098C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RA Gallon L., Dujon B.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z28323; CAA82178.1; -
DR PIR; S38177; S38177.
DR GERMOnline; 140077; -.
DR MEROPS; C19.0PWF; -.
DR SGD; S0001806; UBPI1.
DR GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS02335; UCH_2_3; 1.
KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT DOMAIN 557 564 POLY-HIS.
FT ACT_SITE 307 307 BY SIMILARITY.
FT ACT_SITE 641 641 BY SIMILARITY.
FT ACT_SITE 649 649 BY SIMILARITY.
SQ SEQUENCE 717 AA; 82702 MW; 9BAA1EEB5DD9A65 CRC64;
Query Match 77.6%; Score 38; DB 1; Length 717;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 STAPPVHNV 9
Db 231 ATAPPVHSL 239
[1]
RESULT 4
TRXB MYCTU STANDARD; PRT; 335 AA.
ID TRXB MYCTU
AC P52214; O53592;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Thioredoxin reductase (EC 1.8.1.9) (TRXR) (TR).
GN TRXB OR RV3913 OR MT4032 OR MTV028.04.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
```


RP SEQUENCE FROM N.A.
RA Wieses B., Phillip W., Drifflout J.W., Offringa R.,
RA Ottenhoff T.H.M.;
RT "Sequence analysis and functional characterization of thioredoxin and
RT thioredoxin reductase of Mycobacterium tuberculosis.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickley E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:15479-5490 (2002).
RL J. Bacteriol. 184:15479-5490 (2002).
CC -|- CATALYTIC ACTIVITY: Thioredoxin + NADP(+) = thioredoxin disulfide
CC + NADPH.
CC -|- COFACTOR: Binds 1 FAD per subunit (By similarity).
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- MISCELLANEOUS: The active site is a redox-active disulfide bond.
CC -|- SIMILARITY: Belongs to class-II pyridine nucleotide-disulfide
CC oxidoreductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X95798; CAA5070.1; -;
CC EMBL; AL021426; CAA16226.1; -;
CC EMBL; AE007194; AAK48397.1; -;
CC FTR; A70851; A70851.
CC HSSP; Q39243; 1VDC.
CC TIGR; MT4032; -;
CC TubercuList; RV3913; -;
CC InterPro; IPR000759; Adrndx_reductase.
CC InterPro; IPR001327; FAD_pyr_redux.
CC InterPro; IPR001100; Pyr_redux.
CC InterPro; IPR008255; Pyr_redux2_AS.
CC InterPro; IPR001103; Pyridine_redux_2.
CC InterPro; IPR005982; Thioredux_reduct.
CC Pfam; PF00070; Pyr_redux; 1.
CC PRINTS; PR00419; ADXRDXTASE.
CC PRINTS; PR00368; FADPNR.
CC PRINTS; PR00411; ENDRDTASBI.
CC PRINTS; PR00469; ENDRDTASBI.
CC TIGRams; TTGR01292; TRX_reduct; 1.
CC PROSITE; PS00573; PYRIDINE_REDOX_2; 1.
CC Redox-active center; Oxidoreductase; NADP; Flavoprotein; FAD;
CC Complete proteome.
CC NP_BIND 44 51 FAD (ADP PART) (BY SIMILARITY).
CC FT

FT DISULFID 145 148 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 288 297 FAD (FLAVIN PART) (BY SIMILARITY).
FT CONFLICT 125 125 A -> R (IN REF. 1).
FT CONFLICT 215 215 V -> C (IN REF. 1).
FT CONFLICT 228 228 V -> Y (IN REF. 1).
SQ SEQUENCE 335 AA; 35643 MW; 3D0DD581B6C187E2 CRC64;
Query Match 73.5%; Score 36; DB 1; Length 335;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 TAPPVHN 8
DB 2 TAPPVHD 8
RESULT 5
AFT1 YEAST STANDARD; PRT; 690 AA.
ID AFT1 YEAST
AC P22149;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Iron-regulated transcriptional repressor AFT1.
GN AFT1 OR RCS1 OR YGL071W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95237204; PubMed=7720713;
RA Yamauchi-Iwai Y., Dancis A., Klausner R.D.;
RT "AFT1: a mediator of iron regulated transcriptional control in
RT Saccharomyces cerevisiae.";
RL EMBO J. 14:1231-1239 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Bruckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090 (1997).
RN [3]
RP SEQUENCE OF 283-690 FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91213643; PubMed=2021081;
RA Gil R., Zueco J., Sentandreu R., Herrero E.;
RT "RCS1, a gene involved in controlling cell size in Saccharomyces
RT cerevisiae";
RL Yeast 7:1-14 (1991).
CC -|- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT ACTIVATES THE
CC GENES FOR FRE1, FRE2 AND FRE3 IN RESPONSE TO IRON DEPRIVATION.
CC IRON COULD INTERACT DIRECTLY WITH AFT1 AND INHIBITS ITS ACTIVITY.
CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48004; CAA88044.1; -;
CC EMBL; Z72593; CAA36775.1; -;
CC EMBL; X53046; CAA37215.1; -;
CC PIR; S54775; S54775.
CC GenOnline; 141119; -;
CC TRANSFAC; T03538; -;
CC SGD; S0003039; RCS1.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC DR

DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0006827; P:high affinity iron ion transport; IMP.
DR GO; GO:0045944; P:positive regulation of transcription from P...; IMP.
KW Transcription regulation; Activator; Nuclear protein; Metal-binding;
Iron.
FT VARIANT 291 291 C -> F (IN ALLELE AFT1-LUP; WHICH IS
CONSTITUTIVELY ACTIVATED).
FT CONFLICT 329 329 S -> L (IN REF. 3).
FT CONFLICT 378 378 S -> L (IN REF. 3).
FT CONFLICT 416 416 D -> G (IN REF. 3).
FT CONFLICT 468 468 N -> T (IN REF. 3).
FT CONFLICT 507 507 N -> S (IN REF. 3).
FT CONFLICT 538 538 S -> L (IN REF. 3).
FT CONFLICT 568 568 S -> P (IN REF. 3).
FT CONFLICT 579 579 S -> T (IN REF. 3).
FT CONFLICT 604 644 QHVDHRLSSNEPOVRSHQYGPQOPQOQLQYHQNQPHDGH
-> HMLTIVISALMNLKYDHNVMNRRSSRNCNITKISPT
TGI (IN REF. 3).
FT CONFLICT 649 649 H -> L (IN REF. 3).
FT SEQUENCE 690 AA; 77682 MW; 38641F26B76FCBDD CRC64;
Query Match 73.5%; Score 36; DB 1; Length 690;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 STAPPVHN 8
Db 497 SSAPPVHH 504
RESULT 6
MUTT_STRAM STANDARD; PRT; 154 AA.
AC P32091;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mut-like protein (ORF154).
OS Streptomyces ambifaciens.
OG Plasmid pSAM2.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1889;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 23877;
RC MEDLINE=95020551; PubMed=7934842;
RA Hagege J., Pernodet J.L., Friedmann A., Guerinneau M.;
RT "Mode and origin of replication of pSAM2, a conjugative integrating
element of Streptomyces ambifaciens.";
RL Mol. Microbiol. 10:799-812(1993).
CC -!- SIMILARITY: Belongs to the NUDIX hydrolase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; Z19590; CAA79638.1; -.
DR PIR; S39873; S39873.
DR HSSP; P08337; 1TUM.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR Pfam; PF00293; NUDIX; 1.
DR PRINTS; PR00502; NUDIXFAMILY.
DR PROSITE; PS00893; NUDIX; 1.
KW Plasmid; Hydrolase.
FT DOMAIN 48 69 NUDIX BOX.
FT SEQUENCE 154 AA; 16689 MW; 8EC2EE99D272D656 CRC64;
Query Match 71.4%; Score 35; DB 1; Length 154;

Best Local Similarity 55.6%; Pred. No. 12;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 STAPPVHN 9
Db 11 ATSPPLHSV 19
RESULT 7
LH61_HUMAN STANDARD; PRT; 363 AA.
AC Q9UPM6; Q9NTH3; Q9UPM5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE LIM/homeobox protein Lhx6.1 (Lhx6).
GN LHX6 OR LHX6.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99321857; PubMed=10393337;
RA Kimura N., Ueno M., Nakashima K., Taga T.;
RT "A brain region-specific gene product Lhx6.1 interacts with Ldb1
through tandem LIM-domains.";
RL J. Biochem. 126:180-187(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Amigdala;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansgorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A; Synonyms=Lhx6.1A;
CC IsoId=Q9UPM6-1; Sequence=Displayed;
CC Name=B; Synonyms=Lhx6.1B;
CC IsoId=Q9UPM6-2; Sequence=VSP_003109;
CC -!- TISSUE SPECIFICITY: Brain specific.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AB031041; BAA83422.1; -.
DR EMBL; AB031042; BAA83423.1; -.
DR EMBL; AL136570; CAB66505.1; -.
DR PIR; T46907; T46907.
DR HSSP; P06601; 1FJL.
DR Genew; HGNC:21735; LHX6.
DR MIM; 608215; -.
DR TRANSFAC; T04189; -.
DR TRANSFAC; T04190; -.
DR GO; GO:003700; F:transcription factor activity; TAS.
DR GO; GO:0007420; P:brain development; TAS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR007107; LIM_homeo.

DR	HSPF; P06601; 1FUJ.
DR	TRANSFAC; T04187; -
DR	TRANSFAC; T04188; -
DR	InterPro; IPR001356; Homeobox.
DR	InterPro; IPR001781; LIM.
DR	InterPro; IPR007107; LIM homeo.
DR	Pfam; PF00046; homeobox; 1.
DR	Pfam; PF00412; LIM; 2.
DR	ProDom; PD000010; Homeobox; 1.
DR	ProDom; PD000094; LIM; 2.
DR	SMART; SM00389; HOX; 1.
DR	SMART; SM00132; LIM; 2.
DR	PROSITE; PS00478; LIM DOMAIN 1; 2.
DR	PROSITE; PS00023; LIM DOMAIN 2; 2.
DR	PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
DR	PROSITE; PS00071; HOMEBOX_2; 1.
DR	Homeobox; DNA-binding; Nuclear protein; Repeat; LIM domain;
KW	Metal-binding; Zinc; Transcription regulation; Alternative splicing.
KW	DOMAIN 70 122 LIM 1.
FT	DOMAIN 131 184 LIM 2.
FT	DNA BIND 219 278 HOMEBOX.
FT	VARSPLIC 324 363
FT	
FT	
FT	/FTID=VSP_003110.
SQ	SEQUENCE 363 AA; 40079 MW; F8432368C48EF8BF CRC64;
	Query Match 71.4%; Score 35; DB 1; Length 363;
	Best Local Similarity 100.0%; Pred. No. 28;
	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	2 TAPPVV 7
DB	337 TAPPVV 342
RESULT 9	
PO21_XENLA	
ID ID PO21_XENLA STANDARD; PRT; 760 AA.	
AC Pi6143;	
DT 01-APR-1990 (Rel. 14, Created)	
DT 01-AUG-1991 (Rel. 19, Last sequence update)	
DT 28-FEB-2003 (Rel. 41, Last annotation update)	
DE POU domain, class 2, transcription factor 1 (Octamer-binding	
DE transcription factor 1) (Oct-1) (OTF-1) (NF-A1) (XOctl).	
DE POU2F1 OR OCT1.	
OS Xenopus laevis (African clawed frog).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;	
OC Xenopodinae; Xenopus.	
NCBI_TaxID=8355;	
RN NCBI	
RN RN	
RP SEQUENCE FROM N.A.	
RX MEDLINE=90221827; PubMed=2326173;	
RA Smith D.P., Old R.W.;	
RL "Nucleotide sequence of Xenopus laevis Oct-1 cDNA.";	
RL Nucleic Acids Res. 18:369-369(1990).	
[2]	
RP SEQUENCE OF 261-335 FROM N.A.	
RP Schithuis J.G., Baarends W.M., Peterson-Maduro J., Destre O.H.J.;	
RA Submitted (Feb-1990) to the EMBL/GenBank/DBJ databases.	
RL	
CC -!- FUNCTION: This protein is a transcription factor for small nuclear	
CC RNA and histone H2B genes.	
CC -!- SUBCELLULAR LOCATION: Nuclear.	
CC -!- SIMILARITY: Belongs to the POU transcription factor family.	
CC Class-2 subfamily.	
CC -!- SIMILARITY: Contains 1 homeobox domain.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and/or commercial	


```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17190; CAA35051.1; -.
CC EMBL; X51819; CAA36119.1; ALT_INIT.
CC PIR; S07896; S07896.
CC HSSP; P14859; IOCT.
CC TRANSFAC; T00642; -.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR000972; Octamer-bind_TF.
CC InterPro; IPR000327; POU domain.
CC InterPro; IPR007103; POU_homeo.
CC Pfam; PF00046; homeobox; 1.
CC Pfam; PF00157; pou; 1.
CC PRINTS; PR00029; OCTAMER.
CC PRINTS; PR00028; POUDOMAIN.
CC ProDom; PD000010; Homeobox; 1.
CC ProDom; PD000583; POU domain; 1.
CC SMART; SM00389; Hox; 1.
CC SMART; SM00352; POU; 1.
CC PROSITE; PS00027; HOMEBOX 1; 1.
CC PROSITE; PS00035; POU 1; 1.
CC PROSITE; PS00465; POU_2; 1.
CC PROSITE; PS00071; HOMEBOX 2; 1.
CC Homeobox; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein.
FT DNA_BIND 298 368 POU.
FT DOMAIN 397 456 HOMEBOX.
SQ SEQUENCE 760 AA; 79097 MW; 7066369DA9018345 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 760;
Best Local Similarity 77.8%; Pred.No. 61;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
Db 529 STAPPVSV 537
|||||:|
|

RESULT 10
FINC_XENLA STANDARD; PRT; 2481 AA.
AC Q91740;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fibronectin precursor.
GN FN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92111942; PubMed=1730390;
RA Desimone D.W., Norton P.A., Hynes R.O.;
RT "Identification and characterization of alternatively spliced
RT fibronectin mRNAs expressed in early Xenopus embryos.";
RL Dev. Biol. 149:357-369(1992).
CC -!- FUNCTION: Fibronectins bind cell surfaces and various compounds
CC including collagen, fibrin, heparin, DNA, and actin. Fibronectins
CC are involved in cell adhesion, cell motility, opsonization, wound
CC healing, and maintenance of cell shape (By similarity).
CC -!- SUBUNIT: DIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED VARIANTS,
CC CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS (BY
CC SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. Each of the "extra
CC domain" and the connecting strand 3 are present in some forms of
CC fibronectin and absent in others;
CC Name=1;

```

```

CC -----
CC IsoId=Q91740-1; Sequence=Displayed;
CC TISSUE SPECIFICITY: IN EARLY XENOPUS EMBRYO, CELLULAR FORMS OF
CC FIBRONECTIN PREDOMINATE WHICH INCLUDE BOTH EXTRA DOMAINS. IN
CC FIBRONECTIN OF EMBRYONIC AND ADULT LIVER THE CONNECTING STRAND 3
CC CAN BE EITHER COMPLETELY EXCLUDED OR INCLUDED.
CC -!- SIMILARITY: Contains 12 fibronectin type I domains.
CC -!- SIMILARITY: Contains 2 fibronectin type II domains.
CC -!- SIMILARITY: Contains 17 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77820; AAA49707.1; -.
CC HSSP; P02751; 2FN2.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR000083; Fibinctnl.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR000562; FN_Type_II.
CC Pfam; PF00039; fn1; 12.
CC Pfam; PF00040; fn2; 2.
CC Pfam; PF00041; fn3; 17.
CC PRINTS; PR00013; FNTYPEII.
CC ProDom; PD000995; FN_Type_II; 2.
CC SMART; SM00058; FN1; 12.
CC SMART; SM00059; FN2; 2.
CC SMART; SM00060; FN3; 17.
CC PROSITE; PS00022; EGF 1; 2.
CC PROSITE; PS01253; FIBRONECTIN 1; 11.
CC Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;
KW Repeat; Alternative splicing; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 2481 FIBRONECTIN.
FT DOMAIN 55 275 FIBRIN- AND HEPARIN-BINDING 1.
FT DOMAIN 309 609 COLLAGEN-BINDING.
FT DNA_BIND 907 1172 BY SIMILARITY.
FT DOMAIN 1358 1631 CELL-ATTACHMENT.
FT DOMAIN 1812 2082 HEPARIN-BINDING 2.
FT DOMAIN 2301 2432 FIBRIN-BINDING 2.
FT DOMAIN 53 93 FIBRONECTIN TYPE-I 1.
FT DOMAIN 98 141 FIBRONECTIN TYPE-I 2.
FT DOMAIN 142 185 FIBRONECTIN TYPE-I 3.
FT DOMAIN 187 231 FIBRONECTIN TYPE-I 4.
FT DOMAIN 232 276 FIBRONECTIN TYPE-I 5.
FT DOMAIN 307 346 FIBRONECTIN TYPE-I 6.
FT DOMAIN 346 405 FIBRONECTIN TYPE-II 1.
FT DOMAIN 406 470 FIBRONECTIN TYPE-II 2.
FT DOMAIN 469 512 FIBRONECTIN TYPE-I 7.
FT DOMAIN 517 559 FIBRONECTIN TYPE-I 8.
FT DOMAIN 560 603 FIBRONECTIN TYPE-I 9.
FT DOMAIN 610 707 FIBRONECTIN TYPE-III 1.
FT DOMAIN 708 809 FIBRONECTIN TYPE-III 2.
FT DOMAIN 810 904 FIBRONECTIN TYPE-III 3.
FT DOMAIN 905 995 FIBRONECTIN TYPE-III 4.
FT DOMAIN 996 1085 FIBRONECTIN TYPE-III 5.
FT DOMAIN 1086 1173 FIBRONECTIN TYPE-III 6.
FT DOMAIN 1174 1265 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1266 1356 FIBRONECTIN TYPE-III 8 (EXTRA DOMAIN).
FT DOMAIN 1357 1447 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1448 1537 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1538 1631 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1632 1721 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1722 1811 FIBRONECTIN TYPE-III 13 (EXTRA DOMAIN).
FT DOMAIN 1812 1903 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1904 1992 FIBRONECTIN TYPE-III 15.
FT DOMAIN 1993 2082 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2083 2205 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 2206 2287 FIBRONECTIN TYPE-III 17.

```



```

FT DOMAIN 2299 2343 FIBONECTIN TYPE-I 10.
FT DOMAIN 2344 2386 FIBONECTIN TYPE-I 11.
FT DOMAIN 2388 2431 FIBONECTIN TYPE-I 12.
FT SITE 1615 1617 CELL ATTACHMENT SITE.
FT DISULFID 55 81 BY SIMILARITY.
FT DISULFID 79 90 BY SIMILARITY.
FT DISULFID 100 128 BY SIMILARITY.
FT DISULFID 126 138 BY SIMILARITY.
FT DISULFID 144 172 BY SIMILARITY.
FT DISULFID 170 182 BY SIMILARITY.
FT DISULFID 189 218 BY SIMILARITY.
FT DISULFID 216 228 BY SIMILARITY.
FT DISULFID 234 263 BY SIMILARITY.
FT DISULFID 261 273 BY SIMILARITY.
FT DISULFID 309 336 BY SIMILARITY.
FT DISULFID 334 343 BY SIMILARITY.
FT DISULFID 361 387 BY SIMILARITY.
FT DISULFID 375 402 BY SIMILARITY.
FT DISULFID 421 447 BY SIMILARITY.
FT DISULFID 471 499 BY SIMILARITY.
FT DISULFID 497 509 BY SIMILARITY.
FT DISULFID 519 546 BY SIMILARITY.
FT DISULFID 544 556 BY SIMILARITY.
FT DISULFID 562 590 BY SIMILARITY.
FT DISULFID 588 600 BY SIMILARITY.
FT DISULFID 2301 2330 BY SIMILARITY.
FT DISULFID 2328 2340 BY SIMILARITY.
FT DISULFID 2346 2373 BY SIMILARITY.
FT DISULFID 2371 2383 BY SIMILARITY.
FT DISULFID 2390 2414 BY SIMILARITY.
FT DISULFID 2412 2428 BY SIMILARITY.
FT DISULFID 2429 2459 INTERCHAIN (WITH C-2463) (BY SIMILARITY).
FT DISULFID 2463 2463 INTERCHAIN (WITH C-2459) (BY SIMILARITY).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 543 543 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 877 877 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1291 1291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2202 2202 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2481 AA; 272678 MW; 7E47DF4F6CE72C93 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 2481;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 APPVHV 9
Db 813 APPEHV 819

RESULT 11
NC03 XENLA STANDARD; PRT; 1391 AA.
AC O57539;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear receptor coactivator 3 (EC 2.3.1.48) (Retinoid X receptor-
DE interacting coactivator xSRC-3).
GN NCOA3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A., INTERACTION WITH RXRA; THRA AND EP300, AND
RP MUTAGENESIS OF LEU-622; LEU-683 AND LEU-739.
RC TISSUE=Oocyte;
RX MEDLINE=98322543; PubMed=9658407;
RA Kim H.-J., Lee S.-K., Na S.-Y., Choi H.-S., Lee J.W.;
RA "Molecular cloning of xSRC-3, a novel transcription coactivator from
RT

```

```

RT Xenopus, that is related to AIB1, p/CIP and TIF2.";
RL Mol. Endocrinol. 12:1038-1047(1998).
CC -!- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
CC receptors and stimulates the transcriptional activities in a
CC hormone-dependent fashion. Plays a central role in creating a
CC multisubunit coactivator complex, probably via remodeling of
CC chromatin. Involved in the coactivation of different nuclear
CC receptors, such as retinoids (RAR and RXR), thyroid hormone (TR)
CC and orphan nuclear receptor (hepatocyte nuclear receptor 4 (HNF4)
CC and constitutive androstane receptor (CAR)). Displays histone
CC acetyltransferase activity.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + histone = CoA + acetyl-histone.
CC -!- SUBUNIT: Interacts with the histone acetyltransferase protein
CC EP300.
CC -!- SUBCELLULAR LOCATION: Mainly cytoplasmic and weakly nuclear (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms may be produced;
CC Name=1;
CC IsoId=O57539-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Highly expressed in liver and in early stages
CC of oocyte development.
CC -!- DEVELOPMENTAL STAGE: Expressed only in early stages of oocyte
CC development. Expression is more prominent in stage I, strongly
CC decreases in stage II and then, gradually disappears.
CC -!- DOMAIN: Contains three Leu-Xaa-Xaa-Leu (LXXLL) motifs. Motifs
CC 1 and 2 are essential for the association with nuclear receptors,
CC and constitute the RID domain (Receptor-interacting domain).
CC -!- PTM: Phosphorylated and acetylated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SRC/P160 FAMILY OF NUCLEAR RECEPTOR
CC COACTIVATORS.
CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC -!- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF044080; AAC12927.1; --
CC InterPro; IPR001092; HLH basic.
CC InterPro; IPR000014; PAS domain.
CC InterPro; IPR008955; Src-1.
CC Pfam; PF00010; HLH; 1.
CC Pfam; PF00989; PAS; 1.
CC SMART; SM00353; HLH; 1.
CC SMART; SM00091; PAS; 1.
CC PROSITE; PS50888; HLH; 1.
CC PROSITE; PS50112; PAS; 1.
CC Transferrase; Transcription regulation; Activator; Nuclear protein;
CC Acetylation; Phosphorylation; Repeat; Alternative splicing.
CC DNA_BIND 37 45 BASIC DOMAIN.
CC DOMAIN 46 82 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC DOMAIN 112 182 PAS
CC DOMAIN 1088 1274 ACETYLTRANSFERASE ACTIVITY
CC (BY SIMILARITY).
CC DOMAIN 515 522 POLY-SER.
CC DOMAIN 503 666 SER-RICH.
CC DOMAIN 968 971 POLY-GLN.
CC DOMAIN 1241 1248 POLY-GLN.
CC SITE 680 684 LXXLL MOTIF 1.
CC SITE 736 740 LXXLL MOTIF 2.
CC SITE 1048 1052 LXXLL MOTIF 3.
CC MOD_RES 614 614 ACETYLATION (BY SIMILARITY).
CC MOD_RES 617 617 ACETYLATION (BY SIMILARITY).
CC MOD_RES 618 618 ACETYLATION (BY SIMILARITY).
CC MUTAGEN 622 622 L->A: WEAKLY IMPAIRS INTERACTION WITH
CC NUCLEAR RECEPTORS
CC MUTAGEN 683 683 L->A: STRONGLY IMPAIRS INTERACTION WITH

```



```

FT  MUTAGEN      739      739      NUCLEAR RECEPTORS.
FT  L-SA: STRONGLY IMPAIRS INTERACTION WITH
FT  NUCLEAR RECEPTORS.
SQ  SEQUENCE      1391 AA; 152532 MW; AD28F5CD934AC33D CRC64;

Query Match      67.3%; Score 33; DB 1; Length 1391;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY  1 STAPPVHNV 9
    : |||||
    : |||||
Db   860 SSTPPVENV 868

RESULT 12
POLR_TYMYVA
ID  POLR_TYMYVA STANDARD; PRT; 1844 AA.
AC  P20128;
DT  01-FEB-1991 (Rel. 17, Created)
DT  01-FEB-1991 (Rel. 17, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  RNA replicase polyprotein (EC 2.7.7.48).
OS  Turnip yellow mosaic virus (Austrian isolate).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;
OC  Tymovirus.
OX  NCBI_TaxID=12155;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=90021184; PubMed=2800335;
RA  Keese P., Mackenzie A., Gibbs A.;
RT  "Nucleotide sequence of the genome of an Australian isolate of turnip
RT  yellow mosaic tymovirus."
RL  Virology 172:536-546(1989).
CC  -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC  {RNA}(N).
CC  -|- MISCELLANEOUS: THE 206 kDa PROTEIN IS POTENTIALLY A POLYPROTEIN
CC  (BY HOMOLOGY TO THE LONGER PROTEIN OF THE SINDBIS VIRUS).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
CC  -----
EMBL; J04373; AAA46592.1; -.
PIR; JQ0109; RRWPTM.
InterPro; IPR008043; Peptidase C21.
InterPro; IPR001788; RNA_dep_RNApol2.
InterPro; IPR007095; RNA_pol_DS_PS.
InterPro; IPR007094; RNA_pol_PSVir.
InterPro; IPR006060; Viral_helicase1.
Pfam; PF05381; Peptidase C21; 1.
Pfam; PF00978; RNA_dep_RNApol2; 1.
Pfam; PF01443; Viral_helicase1; 1.
Transferase; RNA-directed RNA polymerase; Polyprotein; ATP-binding.
FT  NP_BIND      976      983      ATP (BY SIMILARITY).
FT  DOMAIN      1631      1665      POLYMERASE SITE (BY SIMILARITY).
SQ  SEQUENCE      1844 AA; 206510 MW; CB447BF0F199A18 CRC64;

Query Match      67.3%; Score 33; DB 1; Length 1844;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  1 STAPPVHNV 8
    : |||||
    : |||||
Db   581 STAPPDHS 588

RESULT 13
PR33 DAUCA
ID  PR33_DAUCA STANDARD; PRT; 211 AA.

```

```

AC  P06600;
DT  01-JAN-1988 (Rel. 06, Created)
DT  01-JAN-1988 (Rel. 06, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Proline-rich 33 kDa extensin-related protein precursor (Fragment).
OS  Daucus carota (Carrot).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC  campanulids; Apiales; Apiaceae; Scandiceae; Daucinae;
OC  Daucus.
OX  NCBI_TaxID=4039;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Chen J., Varner J.E.;
RT  "Isolation and characterization of cDNA clones for carrot extensin
RT  and a proline-rich 33-kDa protein."
RL  Proc. Natl. Acad. Sci. U.S.A. 82:4399-4403(1985).
CC  -|- SIMILARITY: BELONGS TO THE ENOD12 FAMILY OF EXTENSINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
CC  -----
EMBL; M11222; AAA33138.1; -.
PIR; B23162; B23162.
Cell wall; Repeat; Signal.
NON_TER 1 1
FT  SIGNAL      <1      14      POTENTIAL.
FT  CHAIN       15      211      PROLINE-RICH 33 KDA EXTENSIN-RELATED
FT  VARIANT     29      29      PROTEIN.
FT  VARIANT     32      32      T -> K.
FT  VARIANT     32      32      V -> I.
SQ  SEQUENCE     211 AA; 23521 MW; 7897E854C0B48A6E CRC64;

Query Match      65.3%; Score 32; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4 PPVHN 8
    : |||||
    : |||||
Db   110 PPVHN 114

RESULT 14
HI_DROHY
ID  HI_DROHY STANDARD; PRT; 249 AA.
AC  P17268;
DT  01-AUG-1990 (Rel. 15, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Histone H1.
OS  Drosophila hydei (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Aphidrodidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7224;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=90221896; PubMed=2109309;
RA  Kremer H., Hennig W.;
RT  "Isolation and characterization of a Drosophila hydei histone DNA
RT  repeat unit."
RL  Nucleic Acids Res. 18:1573-1580(1990).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Strausbaugh L.D., Fitch D.H.A., Barrett V.;
RA  Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
CC  -|- FUNCTION: Histones H1 are necessary for the condensation of

```



```
CC      nucleosome chains into higher order structures.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- SIMILARITY: Belongs to the histone H1/H5 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X17072; CAA34918.1; -.
CC      EMBL; X52576; CAA36804.1; -.
CC      PIR; S21935; S21935.
CC      HSP; P02259; IHST.
CC      FlyBase; FBgn012374; Dhyd\Hisi.
CC      InterPro; IPR005818; Histone_H1/H5.
CC      InterPro; IPR005819; Histone_H5.
CC      Pfam; PF00538; linker_histone; 1.
CC      PRINTS; PR00624; HISTONEH5.
CC      SMART; SM00526; H15; 1.
CC      Chromosomal protein; Nuclear protein; DNA-binding.
CC      KW      Chromosomal protein; Nuclear protein; DNA-binding.
CC      FT      CONFLICT 208 208 A -> P (IN REF. 2).
CC      ST      SEQUENCE 249 AA; 25746 MW; 189236C05118F903 CRC64;
CC      -----
CC      Query Match      65.3%; Score 32; DB 1; Length 249;
CC      Best Local Similarity 71.4%; Pred. No. 69;
CC      Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC      -----
QY      1 STAPPVH 7
DB      :|||||
DB      40 TTAPPPTH 46

RESULT 15
COQ4_HUMAN
ID      COQ4_HUMAN STANDARD; PRT; 265 AA.
AC      Q9Y3A0; Q96EW4;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Ubiquinone biosynthesis protein COQ4 homolog (Coenzyme Q biosynthesis
DE      protein 4 homolog) (CGI-92).
DE      COQ4.
GN      COQ4.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20272150; PubMed=10810093;
RA      Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT      "Identification of novel human genes evolutionarily conserved in
RT      Caenorhabditis elegans by comparative proteomics.";
RL      Genome Res. 10:703-713(2000).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      TISSUE=Lung;
RA      MEDLINE=22388257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Boraldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettaman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
```

```
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA      Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA      Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT      "Generation and initial analysis of more than 15,000 full-length
RT      human and mouse cDNA sequences.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC      -!- PATHWAY: Coenzyme Q biosynthesis.
CC      -!- SIMILARITY: Belongs to the COQ4 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF151850; AAD34087.1; -.
CC      EMBL; RC011895; AAH11895.1; -.
CC      Genew; HGNC:19693; COQ4.
CC      InterPro; IPR007715; Coq4.
CC      Pfam; PF05019; Coq4; 1.
CC      Ubiquinone biosynthesis.
CC      KW      CONFLICT 230 230 P -> H (IN REF. 1).
CC      ST      SEQUENCE 265 AA; 29671 MW; 1C2B3851E8D0E480 CRC64;
CC      -----
CC      Query Match      65.3%; Score 32; DB 1; Length 265;
CC      Best Local Similarity 83.3%; Pred. No. 74;
CC      Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC      -----
QY      2 TAPPVH 7
DB      :|||||
DB      255 TAPPVH 260

Search completed: July 28, 2004, 18:09:54
Job time : 15 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:08:33 ; Search time 35 Seconds
(without alignments)
81.133 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriaph: *
17: sp_archeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	83.7	590	16	Q8yk80 anabaena sp
2	40	81.6	553	6	Q9mz11 macaca mula
3	40	81.6	803	5	Q7ysx2 trypanosoma
4	39	79.6	20	4	Q9um18 homo sapien
5	39	79.6	499	10	Q9SEH0 pisum sativ
6	39	79.6	802	13	O42127 xenopus lae
7	39	79.6	2768	5	Q9VC00 drosophila
8	37	75.5	1191	3	O14306 schizosacch
9	36	73.5	96	12	Q9QTW0 tt virus. o
10	36	73.5	108	12	Q9QTY6 tt virus. o
11	36	73.5	115	12	Q9WAY6 tt virus. o
12	36	73.5	127	12	Q8V7B8 tt virus. o
13	36	73.5	128	12	Q8V7D7 tt virus. o
14	36	73.5	128	12	Q8V7F1 tt virus. o
15	36	73.5	128	12	Q8V7D5 tt virus. o
16	36	73.5	128	12	Q9JG74 tt virus. o

17	36	73.5	128	12	Q8V7E1
18	36	73.5	130	12	Q8UYG2
19	36	73.5	134	12	Q9DUH9
20	36	73.5	147	12	Q99AR4
21	36	73.5	152	12	Q9DDU0
22	36	73.5	154	12	Q99A81
23	36	73.5	156	12	Q9WSW8
24	36	73.5	156	12	Q70805
25	36	73.5	271	12	Q8V7F5
26	36	73.5	280	12	Q8V7F6
27	36	73.5	305	12	Q8QNK4
28	36	73.5	329	10	Q9SUT6
29	36	73.5	335	16	Q7TWC8
30	36	73.5	341	10	Q9FKQ5
31	36	73.5	391	16	Q7U9W5
32	36	73.5	531	5	Q8T8Q4
33	36	73.5	690	3	Q86993
34	36	73.5	716	13	Q7ZVG5
35	36	73.5	720	5	Q23847
36	36	73.5	1013	5	Q967T5
37	36	73.5	4001	5	Q8WRO7
38	36	73.5	4001	5	Q9VCA8
39	36	73.5	4793	3	Q7Z8E4
40	35	71.4	128	12	Q8V7E8
41	35	71.4	144	11	Q7TT08
42	35	71.4	157	16	Q82E38
43	35	71.4	182	5	Q19136
44	35	71.4	235	10	Q9FRY9
45	35	71.4	253	16	Q81B73

ALIGNMENTS

RESULT 1

Q8YK80 ID Q8YK80 PRELIMINARY; PRT; 590 AA.
AC Q8YK80;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein A18037.
GN A18037.
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120gamma.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003603; BAB77367.1; -.
DR PIR; AF2555; AF2555.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0009291; P:unidirectional conjugation; IEA.
DR InterPro; IPR003688; TRAG.
DR Pfam; PF02534; TRAG; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 590 AA; 64933 MW; 7ECA68DF6FE3B469 CRC64;

Query Match 83.7%; Score 41; DB 16; Length 590;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAPPVHV 9

M27


```

SQ SEQUENCE 499 AA; 56635 MW; 08AC7E24B94802B5 CRC64;
Query Match 79.6%; Score 39; DB 10; Length 499;
Best Local Similarity 87.5%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAPPVHN 8
| | | | |
Db 10 STAPPVHN 17

RESULT 6
O42127 PRELIMINARY; PRT; 802 AA.
AC O42127;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE FGF receptor 3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
ON NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hongo I., Kengaku M., Okamoto H.;
RT "Differential employment of FGF signaling system for the embryonic
RT induction."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB007035; BAA22281.1; -
DR HSSP; P11362; 1FGK.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00047; Ig; 3.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00835; IG LIKE; 3.
DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Immunoglobulin domain; Kinase; Receptor; Transferase;
KW Tyrosine-protein kinase.
SQ SEQUENCE 802 AA; 89515 MW; CC5ESDDF3BD25BB3 CRC64;

Query Match 79.6%; Score 39; DB 13; Length 802;
Best Local Similarity 87.5%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAPPVHN 9
| | | | |
Db 396 TAPPVHN 403

RESULT 7
Q9VC00 PRELIMINARY; PRT; 2768 AA.
AC Q9VC00;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG13648 protein.

```

```

GN CG13648.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Miklos G.D.G.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Fezraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jafarizadeh M., Kalush F., Karpis G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003750; AAF56376.1; -
DR FlyBase; FBgn039257; CG13648.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00214; VWC; 4.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS00184; VWFC_2; 1.
SQ SEQUENCE 2768 AA; 294032 MW; CA929A21774E4684 CRC64;

Query Match 79.6%; Score 39; DB 5; Length 2768;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAPPVHN 9
| | | | |
Db 2332 STEPPVHDV 2340

RESULT 8
O14306 PRELIMINARY; PRT; 1191 AA.
AC O14306;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```



```

DE Putative inositol polyphosphate phosphatase.
GN SPAC9G1.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Churcher C.M., Gentles S., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98763; CAB1494.2; -.
DR FJ; T39233; T39233.
DR GeneDB SPombe; SPAC9G1.10c; -.
DR GO; GO:0004437; F:inositol/phosphatidylinositol phosphatase a. . . IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000300; IPPC.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR SMART; SM00128; IPPC; 1.
SQ SEQUENCE 1191 AA; 131274 MW; 7BAB86AD95F3A7E7 CRC64;

Query Match 75.5%; Score 37; DB 3; Length 1191;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
Db 253 STAPPVNI 261

RESULT 9
Q9QTW0 PRELIMINARY; PRT; 96 AA.
AC Q9QTW0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF2 protein (Fragment).
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTV WAM993;
RX MEDLINE=20129505; PubMed=10664406;
RA Mulyanto, Hijikata M., Matsushita M., Ingkokusmo G., Widjaya A.,
RA Sumarsidi D., Kanai K., Ohta Y., Mishiro S.;
RT "TT virus (TTV) genotypes in native and non-native prostitutes of
RT Irian Jaya, Indonesia: implication for non-occupational
RT transmission.";
RL Arch. Virol. 145:63-72(2000).
DR EMBL; AB031733; BAA86196.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
DR NON_TER 96
FT SEQUENCE 96 AA; 11160 MW; 12P657A821FB2871 CRC64;

Query Match 73.5%; Score 36; DB 12; Length 96;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPVHV 9
Db 36 PPVHV 41

RESULT 10
Q9QTY6 PRELIMINARY; PRT; 108 AA.
AC Q9QTY6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

```

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF2 protein (Fragment).
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTV L021;
RX MEDLINE=20129505; PubMed=10664406;
RA Mulyanto, Hijikata M., Matsushita M., Ingkokusmo G., Widjaya A.,
RA Sumarsidi D., Kanai K., Ohta Y., Mishiro S.;
RT "TT virus (TTV) genotypes in native and non-native prostitutes of
RT Irian Jaya, Indonesia: implication for non-occupational
RT transmission.";
RL Arch. Virol. 145:63-72(2000).
DR EMBL; AB031707; BAA86170.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
DR NON_TER 108
FT SEQUENCE 108 AA; 12062 MW; D0C3E628DE29D3F7 CRC64;

Query Match 73.5%; Score 36; DB 12; Length 108;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPVHV 9
Db 33 PPVHV 38

RESULT 11
Q9WAY6 PRELIMINARY; PRT; 115 AA.
AC Q9WAY6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF2 (Fragment).
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=07NaK/2-4G;
RX MEDLINE=99335592; PubMed=10405352;
RA Hijikata M., Takahashi K., Mishiro S.;
RT "Complete circular DNA genome of a TT virus variant (isolate name
RT SANBAN) and 44 partial ORF2 sequences implicating a great degree of
RT diversity beyond genotypes.";
RL Virology 260:17-22(1999).
DR EMBL; AB024347; BAA77414.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
DR NON_TER 115
FT SEQUENCE 115 AA; 12824 MW; E4B2702604E41109 CRC64;

Query Match 73.5%; Score 36; DB 12; Length 115;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPVHV 9
Db 33 PPVHV 38

RESULT 12
Q8V7B8 PRELIMINARY; PRT; 127 AA.
AC Q8V7B8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```


DE ORF2.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21844401; PubMed=11855633;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064632; BAB79407.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
SQ SEQUENCE 127 AA; 13395 MW; 66E858DA565B625D CRC64;

Query Match 73.5%; Score 36; DB 12; Length 127;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPVHNV 9
Db 6 PPVHNV 11

RESULT 13
Q8V7D7
ID Q8V7D7 PRELIMINARY; PRT; 128 AA.
AC Q8V7D7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ORF2.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21844401; PubMed=11855633;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064619; BAB79381.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
SQ SEQUENCE 128 AA; 12982 MW; 10AC68484604473D CRC64;

Query Match 73.5%; Score 36; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPVHNV 9
Db 5 PPVHNV 10

RESULT 14
Q8V7F1
ID Q8V7F1 PRELIMINARY; PRT; 128 AA.
AC Q8V7F1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ORF2.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=21844401; PubMed=11855633;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064610; BAB79363.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
SQ SEQUENCE 128 AA; 13010 MW; 7B91C3D307B93064 CRC64;

Query Match 73.5%; Score 36; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPVHNV 9
Db 5 PPVHNV 10

RESULT 15
Q8V7D5
ID Q8V7D5 PRELIMINARY; PRT; 128 AA.
AC Q8V7D5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ORF2.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21844401; PubMed=11855633;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064620; BAB79383.1; -.
DR InterPro; IPR004118; TT_ORF2.
DR Pfam; PF02957; TT_ORF2; 1.
SQ SEQUENCE 128 AA; 12971 MW; 0759F8CF3F7D0393 CRC64;

Query Match 73.5%; Score 36; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPVHNV 9
Db 5 PPVHNV 10

Search completed: July 28, 2004, 18:10:40
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:09:33 ; Search time 19 Seconds
(without alignments)
24.454 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	1867	2	US-08-479-537A-5
2	49	100.0	1867	3	US-09-083-116-5
3	49	100.0	1867	4	US-09-134-916A-5
4	49	100.0	2035	2	US-08-479-537A-2
5	49	100.0	2035	3	US-09-083-116-2
6	49	100.0	2035	4	US-09-134-916A-2
7	39	79.6	9	1	US-08-787-547-55
8	39	79.6	9	2	US-08-288-059-19
9	39	79.6	9	4	US-09-593-870A-45
10	39	79.6	9	4	US-09-497-232-1
11	39	79.6	16	4	US-09-043-731-19
12	39	79.6	19	1	US-08-099-354-3
13	39	79.6	19	2	US-08-288-059-9
14	39	79.6	20	2	US-08-288-059-1
15	39	79.6	20	2	US-08-288-059-32
16	39	79.6	20	2	US-08-302-516-20
17	39	79.6	20	2	US-08-833-807-1
18	39	79.6	20	3	US-09-339-944-1
19	39	79.6	20	3	US-08-737-896-3
20	39	79.6	20	3	US-09-223-043-1
21	39	79.6	20	3	US-08-134-198E-34
22	39	79.6	20	4	US-09-847-185-20
23	39	79.6	20	4	US-09-593-870A-1
24	39	79.6	20	4	US-09-646-028-40
25	39	79.6	20	4	US-09-497-232-9
26	39	79.6	20	4	US-09-497-232-11
27	39	79.6	20	4	US-09-651-265-1

28	39	79.6	20	4	US-09-000-003A-3	Sequence 3, Appli
29	39	79.6	20	4	US-09-601-729-140	Sequence 140, App
30	39	79.6	20	5	PCT-US96-09951-3	Sequence 3, Appli
31	39	79.6	21	1	US-08-099-354-4	Sequence 4, Appli
32	39	79.6	21	2	US-08-288-059-10	Sequence 10, Appli
33	39	79.6	21	2	US-08-833-807-14	Sequence 14, Appli
34	39	79.6	21	3	US-09-223-043-14	Sequence 14, Appli
35	39	79.6	21	4	US-09-593-870A-23	Sequence 23, Appli
36	39	79.6	25	2	US-08-288-059-28	Sequence 28, Appli
37	39	79.6	25	4	US-09-497-232-5	Sequence 5, Appli
38	39	79.6	28	2	US-08-488-161-9	Sequence 9, Appli
39	39	79.6	28	3	US-09-273-685-9	Sequence 9, Appli
40	39	79.6	28	5	PCT-US95-11934-9	Sequence 9, Appli
41	39	79.6	30	3	US-08-737-896-6	Sequence 6, Appli
42	39	79.6	30	3	US-08-134-198E-13	Sequence 13, Appli
43	39	79.6	30	4	US-09-593-870A-47	Sequence 47, Appli
44	39	79.6	30	5	PCT-US96-09951-6	Sequence 6, Appli
45	39	79.6	32	4	US-09-593-870A-46	Sequence 46, Appli

ALIGNMENTS

RESULT 1

US-08-479-537A-5

; Sequence 5, Application US/08479537A

; Patent No. 5861381

; GENERAL INFORMATION:

; APPLICANT: CHAMBERON, Pierre

; APPLICANT: KIENY, Marie-Paule

; APPLICANT: LATHE, Richard

; APPLICANT: HARUUVENI, Mara

; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/479,537A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 90/13101

; FILING DATE: 23-OCT-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/FR91/00835

; FILING DATE: 23-OCT-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/039,320

; FILING DATE: 04-APR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/403,576

; FILING DATE: 14-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Teskin, Robin L.

; REGISTRATION NUMBER: 35,030

; REFERENCE/DOCKET NUMBER: 017753-025

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-6620

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1867 amino acids

Thu Jul 29 08:21:26 2004

us-10-019-513-1.rai

```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128-1727 /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat i
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."

```

Query Match	100.0%;	Score 49;	DB 2;	Length 1867;
Best Local Similarity	100.0%;	Pred. No. 4;		
Matches	9.	Mismatches	0;	
Conservative				
			0;	Indels 0;
				Gaps 0;

QY 1 STAPPVHV 9
|||
1730 STAPPVHV 1738
pb

RESULT 2
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMON, pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:

```

> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 08/479,537
> FILING DATE:
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: WO PCT/FR91/00835
> FILING DATE: 23-OCT-1991
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/039,320
> FILING DATE: 04-APR-1993
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/403,576
> FILING DATE: 14-MAR-1995
> ATTORNEY/AGENT INFORMATION:
> NAME: Teskin, Robin L.
> REGISTRATION NUMBER: 35,030
> REFERENCE/DOCKET NUMBER: 017753-025
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (703) 836-6620
> TELEFAX: (703) 836-2021
> INFORMATION FOR SEQ ID NO: 5:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1867 amino acids
> TYPE: amino acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: peptide
> FEATURE:
> NAME/KEY: Peptide
> LOCATION: 128..1727 /note= "The amino acids spanning
> OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat i
> OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
> OTHER INFORMATION: repeats varies from 1 to 40."
> FEATURE:
> NAME/KEY: Peptide
> LOCATION: 134
> OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
> OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
> OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
> FEATURE:
> NAME/KEY: Peptide
> LOCATION: 144
> OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
> OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
> OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
> FEATURE:
> NAME/KEY: Peptide
> LOCATION: 147
> OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
> OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
> OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
> FEATURE:
> NAME/KEY: Peptide
> LOCATION: 1..21
> OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
> OTHER INFORMATION: 21 amino acid precursor sequence."
> US-09-083-116-5
> Query Match 100.0%; Score 49; DB 3; Length 1867;
> Best Local Similarity 100.0%; Pred. No. 4;
> Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
> Qy 1 STAPPVHV 9
> Db 1730 STAPPVHV 1738
> RESULT 3
> US-09-134-916A-5
> Sequence 5, Application US/09134916A
> Patent No. 6328956
> GENERAL INFORMATION:
> APPLICANT: CHAMBEON, Pierre

```

Qy 1 STAPPVHV 9
1730 STAPPVHV 1738

RESULT 3
US-09-134-916A-5
; Sequence 5, Application US/09134916A
; Patent No. 6328956
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre

APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1727 constitute a repeated region wherein the repeat
is 20 amino acids, 17 of which are fixed. The number of such
repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
21 amino acid precursor sequence."
US-09-134-916A-5
Query Match 100.0%; Score 49; DB 4; Length 1867;
Best Local Similarity 100.0%; Pred. NO. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 STAPPVHV 9
Db 1730 STAPPVHV 1738
RESULT 4
US-08-479-537A-2
Sequence 2, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:

NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1899 constitute a repeated region wherein the repeat
of 20 amino acids, 17 of which are fixed. The number of such
repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
21 amino acid precursor sequence."
US-08-479-537A-2

Query Match 100.0%; Score 49; DB 2; Length 2035;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
Db 1730 STAPPVHV 1738

RESULT 5
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMBER, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1899 constitute a repeated region wherein the repeat
of 20 amino acids, 17 of which are fixed. The number of such
repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
21 amino acid precursor sequence."
US-09-083-116-2
Query Match 100.0%; Score 49; DB 3; Length 2035;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 STAPPVHV 9
Db 1730 STAPPVHV 1738
RESULT 6
US-09-134-916A-2
Sequence 2, Application US/09134916A
Patent No. 6328956
GENERAL INFORMATION:
APPLICANT: CHAMBER, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US/08/479,537
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-2
Query Match 100.0%; Score 49; DB 4; Length 2035;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 STAPPVHV 9
Db 1730 STAPPVHV 1738
RESULT 7
US-08-787-547-55
; Sequence 55, Application US/08787547
; Patent No. 5783567
; GENERAL INFORMATION:
; APPLICANT: Hedley, Mary Lynne
; APPLICANT: Curley, Joanne M.
; APPLICANT: Langer, Robert S.
; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
; TITLE OF INVENTION: OF NUCLEIC ACID
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,547
; FILING DATE: 22-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 08191/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-787-547-55
Query Match 79.6%; Score 39; DB 1; Length 9;
Best Local Similarity 77.8%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 STAPPVHV 9
Db 1 STAPPVHV 9
RESULT 8
US-08-288-059-19
; Sequence 19, Application US/08288059
; Patent No. 5827666
; GENERAL INFORMATION:

APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELAPO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-288-059-19

Query Match 79.6%; Score 39; DB 2; Length 9;
Best Local Similarity 77.8%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

Qy 1 STAPPVHNV 9
| | | | |
Db 1 STAPPAHGV 9

RESULT 9
US-09-593-870A-45
; Sequence 45, Application US/09593870A
; Patent No. 6548643
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff Allan
; TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their
; FILE REFERENCE: Use in Immunotherapy
; FILE REFERENCE: 2368-McKenzie
; CURRENT APPLICATION NUMBER: US/09/593,870A
; CURRENT FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: 09/223,043
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-593-870A-45

Query Match 79.6%; Score 39; DB 4; Length 9;
Best Local Similarity 77.8%; Pred. No. 3e+05;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 STAPPVHNV 9
| | | | |
Db 1 STAPPAHGV 9

RESULT 10
US-09-497-232-1
; Sequence 1, Application US/09497232
; Patent No. 660012
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; KRANTZ, Mark J.
; REDDISH, Mark A. Michael
; LONGENECKER, B. Michael
; TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/497,232
FILING DATE: 03-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/074,410
FILING DATE: 08-MAY-1998
APPLICATION NUMBER: US 60/045,949
FILING DATE: 08-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 042881/0114
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-497-232-1

Query Match 79.6%; Score 39; DB 4; Length 9;
Best Local Similarity 77.8%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9
| | | | |
Db 1 STAPPAHGV 9

RESULT 11
US-09-043-731-19
; Sequence 19, Application US/09043731A
; Patent No. 6344203
; GENERAL INFORMATION:
; APPLICANT: The Austin Research Institute
; TITLE OF INVENTION: Mimicking Peptides in Cancer Therapy

FILE REFERENCE: CALA-200
CURRENT APPLICATION NUMBER: US/09/043,731A
CURRENT FILING DATE: 1998-06-23
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 16
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: single
OTHER INFORMATION: stranded linear peptide
US-09-043-731-19

Query Match 79.6%; Score 39; DB 4; Length 16;
Best Local Similarity 77.8%; Pred. No. 1.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAPPVHNV 9
|||||
DB 5 STAPPAHGV 13

RESULT 12
US-08-099-354-3
Sequence 3, Application US/08099354
Patent No. 5744144
GENERAL INFORMATION:
APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/099,354
FILING DATE: 30-JUL-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: SIRILLA, GEORGE M.
REGISTRATION NUMBER: 18221
REFERENCE/DOCKET NUMBER: 6137/202246
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3536
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-099-354-3
Query Match 79.6%; Score 39; DB 1; Length 19;
Best Local Similarity 77.8%; Pred. No. 1.6;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAPPVHNV 9
|||||

Db 8 STAPPAHGV 16
RESULT 13
US-08-288-059-9
Sequence 9, Application US/08288059
Patent No. 5827666
GENERAL INFORMATION:
APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSER: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-288-059-9

Query Match 79.6%; Score 39; DB 2; Length 19;
Best Local Similarity 77.8%; Pred. No. 1.6;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAPPVHNV 9
|||||
DB 8 STAPPAHGV 16

RESULT 14
US-08-288-059-1
Sequence 1, Application US/08288059
Patent No. 5827666
GENERAL INFORMATION:
APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSER: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA

Thu Jul 29 08:21:26 2004

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-288-059-1

Query Match 79.6%; Score 39; DB 2; Length 20;
Best Local Similarity 77.8%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 STAPPVHNV 9
Db 9 STAPPAHGV 17
Search completed: July 28, 2004, 18:11:36
Job time : 19 secs

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-288-059-1

RESULT 15
US-08-288-059-32
Sequence 32, Application US/08288059
Patent No. 5827666
GENERAL INFORMATION:
APPLICANT: FINN OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:06:58 ; Search time 52 Seconds
(without alignments)
48.902 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHNV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A: Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	49	100.0	9	AAB11114	Human MUC
2	49	100.0	9	ABG79089	Human MUC
3	49	100.0	9	ADA50588	Mucin 1 (
4	49	100.0	13	AAW77232	Peptide s
5	49	100.0	30	AAU84987	Human MUC
6	49	100.0	173	AAU84987	Human MUC
7	49	100.0	180	AAU84987	Human MUC
8	49	100.0	287	AAU84987	Human MUC
9	49	100.0	295	AAU84987	Human MUC
10	49	100.0	307	AAU84987	Human MUC
11	49	100.0	312	AAU84987	Human MUC
12	49	100.0	327	AAU84987	Human MUC
13	49	100.0	348	AAU84987	Human MUC
14	49	100.0	455	AAU84987	Human MUC
15	49	100.0	455	AAU84987	Human MUC
16	49	100.0	473	AAU84987	Human MUC
17	49	100.0	473	AAU84987	Human MUC
18	49	100.0	475	AAU84987	Human MUC
19	49	100.0	475	AAU84987	Human MUC
20	49	100.0	475	AAU84987	Human MUC
21	49	100.0	475	AAU84987	Human MUC
22	49	100.0	475	AAU84987	Human MUC
23	49	100.0	495	AAU84987	Human MUC
24	49	100.0	508	AAU84987	Human MUC
25	49	100.0	515	ABG96378	Human ova

ALIGNMENTS

RESULT 1

AAB11114
ID AAB11114 standard; peptide; 9 AA.
XX AAB11114;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human MUC-1 protein fragment SEQ ID NO 1.
XX
KW Human; MUC-1; tumor; HLA-A2 restricted immune reaction; treatment;
KW human leukocyte antigen; gene therapy; antigen-presenting cell.
XX
OS Homo sapiens.
XX
PN DE19917195-A1.
XX
PD 19-OCT-2000.
XX
PF 16-APR-1999; 99DE-01017195.
XX
PR 16-APR-1999; 99DE-01017195.
XX (UYTU-) UNIV TUEBINGEN EBERHARD-KARLS.
XX
XX Brossart P, Stevanovic S, Brugger W, Kanz L, Rammensee HG;
XX WPI; 2001-032872/05.
XX
PT New peptide derived from the MUC-1 tumor marker, used to induce a
XX cytotoxic T cell response for treatment or prevention of tumors.
XX
PS Claim 1; Page 6; 8pp; German.
XX
CC This invention describes a novel peptide (I) derived from the MUC-1 gene
CC which is able to induce an HLA (human leukocyte antigen)-A2-restricted
CC immune reaction against tumor cells. (I) or the nucleic acid (II)
CC encoding (I), are used to induce an immune response against tumor cells,
CC so are useful for treatment or prevention of tumors, in conjunction with
CC other tumor therapies. In particular (II) is used in gene therapy or for
CC in vitro transfection or transformation of cells (particularly antigen-
CC presenting cells, optionally in vivo), for expression of (I). (I) has a
CC high binding capacity for HLA-A2 and can reverse the usual suppression of
CC the immune response associated with tumor cells. By introducing the
CC nucleic acid that encodes (I) into an antigen-presenting cell in vitro,
CC then returning the cells to the patient, a more certain and controlled
CC response is achieved, compared with administration of the peptide plus
CC adjuvant

26 49 100.0 515 6 ADA50565
27 49 100.0 515 6 ABR92123 Human cer
28 49 100.0 515 6 AAE37797 Human muc
29 49 100.0 577 3 AAY71030
30 49 100.0 1255 5 ABF56040
31 49 100.0 1255 6 ABR82568
32 49 100.0 1255 6 ABR47537
33 49 100.0 1255 6 ABR92124 Human cer
34 49 100.0 1255 7 ADD45111 Human pro
35 49 100.0 1255 7 ADE54622 Human pro
36 49 100.0 5546 5 AAU85008 Human mel
37 49 91.8 321 3 AAB3416 Human can
38 39 79.6 9 2 AAR68013 Mucin pep
39 39 79.6 9 2 AAW78844 MUC-1 pro
40 39 79.6 9 2 AAW72715 Mucin pep
41 39 79.6 9 2 AAY46704 Immunogen
42 39 79.6 9 2 AAW67583 T-cell ac
43 39 79.6 9 3 AAB33656 MHC class
44 39 79.6 9 4 AAE09540 Human HLA
45 39 79.6 9 4 AAB95902 MHC class

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9
| | | | | | | | |
Db 1 STAPPVHNV 9

RESULT 2
ID ABG79089 standard; peptide; 9 AA.
XX AC ABG79089;
XX DT 15-NOV-2002 (first entry)
XX DE Human MUCI class I HLA widely expressed antigen peptide #2.
XX KW Cell penetrating peptide; cancer; tumour; melanoma; thymoma; antigen;
XX KW lymphoma; sarcoma; lung cancer; non-Hodgkin's lymphoma; leukaemia;
XX KW Hodgkin's lymphoma; uterine cancer; cervical cancer; bladder cancer;
XX KW kidney cancer; adenocarcinoma; breast cancer; prostate cancer;
XX KW ovarian cancer; pancreatic cancer; epitope; vaccine; dendritic cell;
XX KW tumour infiltrating lymphocyte; TIL; human leukocyte antigen; HLA;
XX KW cytostatic; human.
XX OS Homo sapiens.
XX FN WO200264057-A2.
XX PD 22-AUG-2002.
XX PF 15-FEB-2002; 2002WO-US005212.
XX PR 15-FEB-2001; 2001US-0268687P.
XX PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX PI Wang R;
XX DR WPI; 2002-627577/67.
XX PT Novel composition for treating a disease in an animal, comprises an
XX PT immune effector cell and cell penetrating peptide associated with an
XX PT antigen or antibody.
XX PS Disclosure; Page 18; 61pp; English.

XX CC The invention relates to a composition (I) comprising an immune effector
XX CC cell and a cell penetrating peptide (CPP) associated with an antigen or
XX CC antibody. Also included are (1) a vaccine comprising (I), CPP associated
XX CC with an antigen, and a pharmaceutically acceptable carrier and (2)
XX CC preparing a composition for a disease, by providing (1) and CPP
XX CC associated with an antigen for disease, and introducing the antigen-
XX CC associated CPP to (1), where antigen enters into the cell. The antigens
XX CC are, for example, tumour antigen derived epitopes recognised by tumour
XX CC infiltrating lymphocytes (TIL) of HLA (human leukocyte antigen) class I
XX CC or II. The composition is useful for enhancing immunity in an animal to a
XX CC disease, by administering a mature dendritic cell comprising CPP
XX CC associated with an antigen to disease, to the animal, such that following
XX CC the administration, animal is protected from disease, where the animal
XX CC comprises both CD4+ and CD8+ T cells. It is also useful for treating a
XX CC disease (e.g. cancer, tumour, melanoma, thymoma, lymphoma, sarcoma, lung
XX CC cancer, non-Hodgkin's lymphoma, leukaemia, Hodgkin's lymphoma, uterine
XX CC cancer, cervical cancer, bladder cancer, kidney cancer, adenocarcinoma,
XX CC breast cancer, prostate cancer, ovarian cancer and pancreatic cancer).
XX CC The animal is further subjected to a cancer treatment including surgery,
XX CC radiation, chemotherapy or gene therapy. The administration of (I),
XX CC preferably dendritic cell is prior to, subsequent to or concurrent with,

CC the cancer treatment. The present sequence is a tumour antigen derived
CC epitope for inclusion in the composition of the invention

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9
| | | | | | | | |
Db 1 STAPPVHNV 9

RESULT 3
ID ADA50588 standard; peptide; 9 AA.
XX AC ADA50588;
XX DT 20-NOV-2003 (first entry)
XX DE Mucin 1 (MUC-1) CTL epitope, SEQ ID NO:43.
XX KW Nucleic acid vaccine; DNA vaccine; tumour antigen; cytokine adjuvant;
XX KW humoral response; cellular response; immune response; immunotherapy;
XX KW cancer; cytostatic; vaccine; gene therapy; mucin 1; MUC-1;
XX KW cytotoxic T lymphocyte; CTL epitope.
XX OS Unidentified.
XX FN WO2003031569-A2.
XX PD 17-APR-2003.
XX PF 18-SEP-2002; 2002WO-US029640.
XX PR 10-OCT-2001; 2001US-0328371P.
XX PA (CENZ) CENTOCOR INC.
XX PI Snyder L, Scallion B, Knight DM, McCarthy SG, Goletz TJ;
XX PI Branigan PJ;
XX DR WPI; 2003-393437/37.
XX PT New nucleic acid vaccine, useful for eliciting an immune response to a
XX PT cancer associated tumor protein in a mammal.
XX PS Claim 1a; Page 45; 92pp; English.

XX CC The invention relates to a nucleic acid vaccine comprising one or more
XX CC tumour antigen-encoding nucleic acids and one or more cytokine adjuvant-
XX CC encoding nucleic acids. The tumour antigen encoded by the vaccine is
XX CC mucin 1 (MUC-1), the kallikrein KLK2, or prostate specific antigen (PSA,
XX CC also known as KLK3), and the cytokine adjuvant encoded can be interleukin
XX CC -12 (IL-12), granulocyte macrophage-colony stimulating factor (GM-CSF),
XX CC or especially interleukin-18 (IL-18). The antigen-encoding nucleic acid
XX CC is preferably under the control of a promoter such as the cytomegalovirus
XX CC immediate early promoter, the dihydrofolate reductase promoter or the
XX CC early or late SV40 promoters. The invention also encompasses the method
XX CC of eliciting an immune response to a tumour antigen in a mammal using the
XX CC vaccine of the invention. Coexpression of the antigen and adjuvant
XX CC induces a humoral or cellular response to the tumour antigen, generating
XX CC an immune response useful for treatment or prophylaxis of cancers. The
XX CC present sequence represents a mucin 1 (MUC-1) polypeptide sequence which
XX CC is specifically claimed for use in the vaccine of the invention.

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 6; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

QY      1 STAPPVHNV 9
Db      1 STAPPVHNV 9

RESULT 4
AAW77232
ID      AAW77232 standard; peptide; 13 AA.
XX      AC
XX      AAW77232;
XX      20-NOV-1998 (first entry)
XX      Peptide sequence encoding MUC1 tandem repeat unit c.
DE      MUC1; recombinant pox virus; cytotoxic T-lymphocyte; immunogen; tumour;
KW      tumour-associated antigen.
XX      Homo sapiens.
XX      WO9837095-A2.
XX      27-AUG-1998.
XX      24-FEB-1998; 98WO-US003693.
XX      24-FEB-1997; 97US-0038253P.
XX      (THER-) THERION BIOLOGICS CORP.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA      (DAND ) DANA FARBER CANCER INST INC.
XX      Schlom J, Kantor J, Kufe D, Panicali D, Gritz L;
XX      WPI; 1998-467492/40.
XX      New recombinant pox virus for tumour therapy - comprises DNA encoding an
PT      immunogenic mini-MUC1 fragment comprising 5-25 MUC1 tandem repeat units.
XX      Example 1; Page 20; 42pp; English.
XX      The MUC1 tandem repeat units AAW77230-W77232 were used to create an
CC      immunogenic mini-MUC1 fragment for inclusion in a recombinant pox virus
CC      (RPV). The RPV was used in a pharmaceutical composition also containing
CC      an immunomodulator to generate MUC1 specific cytotoxic T-lymphocytes. The
CC      recombinant pox virus therefore encodes an immunogenic MUC1 fragment that
CC      does not undergo significant genetic deletion, thereby providing an
CC      unexpectedly stable and immunogenic pox virus. They can be used to
CC      prevent or treat tumours expressing MUC1 tumour-associated antigens
XX      SQ
XX      Sequence 13 AA;

Query Match      100.0%; Score 49; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 STAPPVHNV 9
Db      2 STAPPVHNV 10

RESULT 5
AAU84987
ID      AAU84987 standard; peptide; 30 AA.
XX      AC
XX      AAU84987;
XX      08-MAY-2002 (first entry)
XX      Human MUC1R segment 1.
XX      Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW

```

```

KW      viral infection; human immunodeficiency virus; melanoma;
KW      bacterial infection; Salmonella; Legionella; parasitic infection;
KW      Trypanosoma; Toxoplasma; Giardia.
XX      Homo sapiens.
XX      WO200190197-A1.
XX      29-NOV-2001.
XX      25-MAY-2001; 2001WO-AU000622.
XX      26-MAY-2000; 2000AU-00007761.
XX      (AUSU ) UNIV AUSTRALIAN NAT.
XX      Thomson SA, Ramshaw IA;
XX      WPI; 2002-147575/19.
XX      N-PSDB; ABK36807.
XX      New synthetic polypeptides having several different segments of at least
PT      one parent polypeptide linked together differently compared to the
PT      linkage in the parent polypeptide, for inducing immune response against a
PT      pathogen or cancer.
XX      Example 3; Fig 27; 364pp; English.
XX      The invention relates to a new synthetic polypeptide (I) comprising
CC      several different segments of at least one parent polypeptide linked
CC      together in a different relationship relative to their linkage in the
CC      parent polypeptide to impede, abrogate or otherwise alter at least one
CC      function associated with the parent polypeptide and for inducing an
CC      immune response against a pathogen or cancer. Also included are a
CC      synthetic polynucleotide encoding and a computer system for designing the
CC      synthetic polypeptides. The synthetic polypeptides and polynucleotides
CC      are referred to as a Savine. The synthetic polypeptide is useful for
CC      modulating immune responses preferably directed against a pathogen or a
CC      cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head
CC      and neck, pancreas, prostate, stomach, bladder, kidney, bone liver,
CC      oesophagus, brain, testicle, uterus), as potentiating agents.
CC      Compositions comprising the polypeptide may be used in the treatment or
CC      prophylaxis against viral (such as infections caused by HIV (human
CC      immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC      virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC      (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
CC      Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
CC      (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
CC      Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is
CC      a peptide derived from a parent protein used to construct a savine of the
CC      invention
XX      SQ
XX      Sequence 30 AA;

Query Match      100.0%; Score 49; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 STAPPVHNV 9
Db      9 STAPPVHNV 17

RESULT 6
AAY71021
ID      AAY71021 standard; protein; 173 AA.
XX      AC
XX      AAY71021;
XX      29-AUG-2000 (first entry)
XX      Human Mucin 1 (MUC-1) protein fragment #2.
XX

```


KW Human; Mucin 1; MUC-1; tumour; pMRS30 expression vector; anti-tumour;
 KW therapy; immune response; cytostatic; vaccine.
 XX
 OS Homo sapiens.
 XX WO200025827-A2.
 XX PD 11-MAY-2000.
 XX PF 18-OCT-1999; 99WO-EP007874.
 XX PR 30-OCT-1998; 98IT-MI002330.
 XX PA (MENA) MENARINI RICERCH SPA.
 XX PI Parente D, Di Massimo AM, De Santis R;
 XX DR WPI; 2000-365410/31.
 XX N-PSDB; AAD00385.
 XX Composition containing one or more DNA molecules encoding fragments of a
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 PT tumor therapy.
 XX PS Claim 16; Fig 2; 56pp; English.
 XX CC The present sequence is a fragment of human Mucin 1 (MUC-1), an antigenic
 CC protein overexpressed in tumour cells. The sequence was obtained from
 CC B720 tumour cells. The corresponding DNA sequence is cloned into a pMRS30
 CC expression vector and used in pharmaceutical composition e.g. vaccine for
 CC inducing an antigen-specific anti-tumour immune response. Composition
 CC containing this DNA molecule is useful in anti-tumour therapy of patients
 CC affected with tumours characterised by high MUC-1 expression
 XX SQ Sequence 173 AA;
 Query Match 100.0%; Score 49; DB 3; Length 173;
 Best Local Similarity 100.0%; Pred. No. 1.5; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0;
 QY 1 STAPPVHNV 9
 DB 127 STAPPVHNV 135
 RESULT 7
 AAR27664
 ID AAR27664 standard; protein; 180 AA.
 XX
 AC AAR27664;
 XX DT 25-MAR-2003 (revised)
 DT 06-NOV-1992 (first entry)
 XX C-terminal region of H23-ETA-S antigen.
 XX
 KW Secreted; human epithelial antigen; Monoclonal antibody H23; vaccine;
 KW malignant tumour; breast cancer; tandem repeat.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH Region 1..40
 FT /note= "contains 2 tandem repeats - can have up to 80
 FT copies"
 FT Misc-difference 7
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT Misc-difference 17
 FT /label= Thr, Asn
 FT /note= "natural polymorphism"
 FT Misc-difference 20
 FT /label= Pro, Ala

FT Misc-difference 27
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT Misc-difference 37
 FT /label= Thr, Asn
 FT /note= "natural polymorphism"
 FT Misc-difference 40
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 XX WO9207000-A1.
 XX PD 30-APR-1992.
 XX PF 23-OCT-1991; 91WO-FR000835.
 XX PR 23-OCT-1990; 90FR-00013101.
 XX (TRGE) TRANSGENE SA.
 XX Chabon P, Kieny MP, Lathe R, Hareuveni M;
 PI WPI; 1992-167097/20.
 DR N-PSDB; AAQ24681.
 XX Compens. contg. polypeptide antigen recognised by antibody H23 - for
 FT treatment of mammary tumours, also for pox virus compsns. for use in
 FT vaccines.
 XX Claim 3; Page 19-21; 29pp; French.
 CC The tumour antigen recognised by antibody H23 is aberrantly expressed in
 CC epithelial cells from cancerous mammary tissue in about 90 per cent of
 CC breast cancer cases; in a normal individual expression is negligible. The
 CC antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S).
 CC Both forms show a high degree of polymorphism. A 20 amino acid subunit in
 CC ETA can be tandemly repeated up to 80 times. (N.B. two tandem repeats are
 CC shown here; the first half of the amino acid sequence, i.e. on the N-
 CC terminal side of the repeat region, is given in AAR27663). From one
 CC subunit to the next, 1 to 3 amino acids can differ. See also AAQ24678-
 CC Q24681, AAQ29276-7 and AAR23974-R23981. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 XX SQ Sequence 180 AA;
 Query Match 100.0%; Score 49; DB 2; Length 180;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHNV 9
 DB 43 STAPPVHNV 51
 RESULT 8
 AAR27665
 ID AAR27665 standard; protein; 287 AA.
 XX
 AC AAR27665;
 XX DT 25-MAR-2003 (revised)
 DT 06-NOV-1992 (first entry)
 XX Secreted form of H23-ETA antigen.
 XX
 KW ETA-S; human epithelial antigen; Monoclonal antibody H23; vaccine;
 KW malignant tumour; breast cancer; tandem repeat.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH Peptide 1..21

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:10:44 ; Search time 42 Seconds
(without alignments)
67.218 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pap.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pap.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	49	100.0	9	12	US-10-247-703-43	Sequence 43, Appl
2	49	100.0	9	16	US-10-447-161-98	Sequence 98, Appl
3	49	100.0	13	14	US-10-057-136-18	Sequence 18, Appl
4	49	100.0	30	12	US-10-296-734-1168	Sequence 1168, Ap
5	49	100.0	307	12	US-10-247-703-26	Sequence 26, Appl
6	49	100.0	312	12	US-10-296-734-824	Sequence 824, Appl
7	49	100.0	475	12	US-10-247-703-22	Sequence 22, Appl
8	49	100.0	475	15	US-10-417-312-1	Sequence 1, Appl
9	49	100.0	508	14	US-10-057-136-20	Sequence 20, Appl
10	49	100.0	515	12	US-10-612-090-19	Sequence 19, Appl
11	49	100.0	515	12	US-10-247-703-20	Sequence 20, Appl
12	49	100.0	515	14	US-10-097-340-212	Sequence 212, Appl
13	49	100.0	515	14	US-10-171-311-156	Sequence 156, Appl
14	49	100.0	1255	10	US-09-996-069-10	Sequence 10, Appl
15	49	100.0	1255	14	US-10-171-311-158	Sequence 158, Appl

16	49	100.0	1255	14	US-10-177-293-311	Sequence 311, Appl
17	49	100.0	5546	12	US-10-296-734-1210	Sequence 1210, Ap
18	45	91.8	321	9	US-09-925-301-861	Sequence 861, Appl
19	39	79.6	9	9	US-09-909-460-55	Sequence 55, Appl
20	39	79.6	9	12	US-10-296-317-44	Sequence 44, Appl
21	39	79.6	9	12	US-10-247-703-45	Sequence 45, Appl
22	39	79.6	9	12	US-09-872-836-55	Sequence 55, Appl
23	39	79.6	9	16	US-10-447-161-97	Sequence 97, Appl
24	39	79.6	12	12	US-10-247-703-47	Sequence 47, Appl
25	39	79.6	12	16	US-10-447-161-138	Sequence 138, Appl
26	39	79.6	20	9	US-09-847-185-20	Sequence 20, Appl
27	39	79.6	20	9	US-09-984-183-11	Sequence 11, Appl
28	39	79.6	20	9	US-09-984-333-1	Sequence 1, Appl
29	39	79.6	20	12	US-10-612-090-3	Sequence 3, Appl
30	39	79.6	20	12	US-09-994-466-1	Sequence 1, Appl
31	39	79.6	20	14	US-10-057-136-1	Sequence 1, Appl
32	39	79.6	20	14	US-10-057-136-16	Sequence 16, Appl
33	39	79.6	20	14	US-10-057-136-17	Sequence 17, Appl
34	39	79.6	20	14	US-10-224-286-20	Sequence 20, Appl
35	39	79.6	20	14	US-10-335-394-40	Sequence 40, Appl
36	39	79.6	20	15	US-10-406-317-31	Sequence 31, Appl
37	39	79.6	20	16	US-10-297-168-31	Sequence 31, Appl
38	39	79.6	20	16	US-10-716-293-215	Sequence 215, Appl
39	39	79.6	20	16	US-10-441-779C-32	Sequence 32, Appl
40	39	79.6	21	14	US-10-062-710-196	Sequence 196, Appl
41	39	79.6	21	14	US-10-062-710-207	Sequence 207, Appl
42	39	79.6	21	16	US-10-380-927-1	Sequence 1, Appl
43	39	79.6	24	12	US-09-815-346-1	Sequence 1, Appl
44	39	79.6	25	12	US-10-296-317-63	Sequence 63, Appl
45	39	79.6	26	12	US-09-815-346-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-10-247-703-43
; Sequence 43, Application US/10247703
; Publication No: US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallan, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; FILE REFERENCE: CEN310
; CURRENT APPLICATION NUMBER: US/10/247,703
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/328,371
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-43

Query Match 100.0%; Score 49; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
Db 1 STAPPVHV 9

RESULT 2
US-10-447-161-98
; Sequence 98, Application US/10447161

GENERAL INFORMATION:
APPLICANT: THOMPSON, SCOTT A
APPLICANT: RAMSDEN, JAY
TITLE OF INVENTION: Synthetic molecules and uses therefor
FILED OFFICE: St. Louis, Missouri
CURRENT APPLICATION NUMBER: US/10/296,734
CURRENT FILING DATE: 2003-08-04
PRIOR APPLICATION NUMBER: AU PQ7761/00
PRIOR FILING DATE: 2000-05-26


```
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 824
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: MUC1R consensus polypeptide
US-10-296-734-824

Query Match      100.0%; Score 49; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy      1 STAPPVHNV 9
Db      7 STAPPVHNV 15

RESULT 7
US-10-247-703-22
; Sequence 22, Application US/10247703
; Publication No. US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallion, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: CYTOKINE ADJUVANT ENCODING NUCLEIC ACID
; FILE REFERENCE: CEN310
; CURRENT APPLICATION NUMBER: US/10/247.703
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/328,371
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-22

Query Match      100.0%; Score 49; DB 12; Length 475;
Best Local Similarity 100.0%; Pred. No. 3.2; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy      1 STAPPVHNV 9
Db      170 STAPPVHNV 178

RESULT 8
US-10-417-312-1
; Sequence 1, Application US/10417312
; Publication No. US20030235868A1
; GENERAL INFORMATION:
; APPLICANT: Dyax Corp
; TITLE OF INVENTION: Antibodies Specific for Mucin Polypeptide
; FILE REFERENCE: 2403/2002
; CURRENT APPLICATION NUMBER: US/10/417,312
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/374,432
; PRIOR FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-417-312-1
```

```
Query Match      100.0%; Score 49; DB 15; Length 475;
Best Local Similarity 100.0%; Pred. No. 3.2; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy      1 STAPPVHNV 9
Db      170 STAPPVHNV 178
```

```
RESULT 9
US-10-057-136-20
; Sequence 20, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFEL, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20
```

```
Query Match      100.0%; Score 49; DB 14; Length 508;
Best Local Similarity 100.0%; Pred. No. 3.4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy      1 STAPPVHNV 9
Db      203 STAPPVHNV 211
```

```
RESULT 10
US-10-612-090-19
; Sequence 19, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Exemplary Muc1 protein
US-10-612-090-19
```

```
Query Match      100.0%; Score 49; DB 12; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
```


QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 11
US-10-247-703-20
; Sequence 20, Application US/10247703
; Publication No. US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallan, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; FILE REFERENCE: CEN310
; CURRENT FILING DATE: 2002-09-20
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-20

Query Match 100.0%; Score 49; DB 12; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 12
US-10-097-340-212
; Sequence 212, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHWANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 13
US-10-171-311-156
; Sequence 156, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-156

Query Match 100.0%; Score 49; DB 14; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 14
US-09-996-069-10
; Sequence 10, Application US/0996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 11
US-10-247-703-20
; Sequence 20, Application US/10247703
; Publication No. US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallan, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; FILE REFERENCE: CEN310
; CURRENT FILING DATE: 2002-09-20
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-20

Query Match 100.0%; Score 49; DB 12; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 12
US-10-097-340-212
; Sequence 212, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHWANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 13
US-10-171-311-156
; Sequence 156, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-156

Query Match 100.0%; Score 49; DB 14; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAPPVHV 9
| | | | | | | | | |
Db 210 STAPPVHV 218

RESULT 14
US-09-996-069-10
; Sequence 10, Application US/0996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia


```

; APPLICANT: Bamdad, K. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10

```

```

Query Match      100.0%; Score 49; DB 10; Length 1255;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 STAPPVHNV 9
Db      950 STAPPVHNV 958

```

```

RESULT 15
US-10-171-311-158
; Sequence 158, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

```

```

Query Match      100.0%; Score 49; DB 14; Length 1255;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 STAPPVHNV 9
Db      950 STAPPVHNV 958

```

```

Search completed: July 28, 2004, 18:16:11
Job time : 42 secs

```


FT Protein /label= signal
 FT 22. .287
 FT /label= ETA-T
 FT Misc-difference 134
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT Misc-difference 144
 FT /label= Thr, Asn
 FT /note= "natural polymorphism"
 FT Misc-difference 147
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT
 FT
 PN WO9207000-A1.
 XX
 XX
 PD 30-APR-1992.
 XX
 XX 23-OCT-1991; 91WO-FR000835.
 XX
 XX 23-OCT-1990; 90FR-00013101.
 XX
 PA (TRGE) TRANSGENE SA.
 XX
 XX Chambon P, Kieny MP, Lathe R, Hareuveni M;
 XX
 XX WPI: 1992-167097/20.
 DR N-PSDB; AAQ29277.
 DR
 XX
 FT Compens. contg. polypeptide antigen recognised by antibody H23 - for
 FT treatment of mammary tumours, also for pox virus compsns. for use in
 FT vaccines.
 PT
 PT
 XX
 XX Claim 3; Page 19-21; 29pp; French.
 XX
 CC The tumour antigen recognised by antibody H23 is aberrantly expressed in
 CC epithelial cells from cancerous mammary tissue in about 90 per cent of
 CC breast cancer cases; in a normal individual expression is negligible. The
 CC antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S).
 CC Both forms show a high degree of polymorphism. A 20 amino acid subunit in
 CC ETA can be tandemly repeated up to 80 times. From one subunit to the
 CC next, 1 to 3 amino acids can differ. DNA coding for immunogenic fragments
 CC of ETA can be inserted into e.g. vaccinia viruses for treatment of
 CC mammary tumours. See also AAQ24678-Q24681, AAQ29276-7 and AAR23974-
 CC R23981. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 287 AA;
 Query Match 100.0%; Score 49; DB 2; Length 287;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHV 9
 DB 150 STAPPVHV 158
 RESULT 9
 AAY71027
 ID AAY71027 standard; protein; 295 AA.
 XX
 XX AAY71027;
 XX
 DT 12-SEP-2003 (revised)
 DT 29-AUG-2000 (first entry)
 XX
 XX Ubiquitin-E. coli LacI-human Mucin 1 fusion protein #2.
 DE
 KW Ubiquitin; LacI; beta-galactosidase; fusion protein; human; Mucin 1;
 KW MUC-1; tumour; pMRS30 expression vector; anti-tumour; therapy;
 KW immune response; cytostatic; vaccine.
 XX
 OS Homo sapiens.
 OS Escherichia coli.

OS Chimeric.
 XX
 XX Key Location/Qualifiers
 FT Region 1. .123
 FT /label= UBILacI_protein
 FT /note= "contains ubiquitin sequence fused to a portion of
 FT E. coli LacI"
 FT 124. .295
 FT /label= Human_MUC-1_fragment
 XX
 PN WO200025827-A2.
 XX
 XX 11-MAY-2000.
 PD
 XX 18-OCT-1999; 99WO-EP007874.
 XX
 XX 30-OCT-1998; 98IT-MI002330.
 DR
 XX (MENA) MENAFINI RICERCHE SPA.
 PA
 XX Parente D, Di Massimo AM, De Santis R;
 PI
 XX WPI: 2000-365410/31.
 DR N-PSDB; AAD00391.
 DR
 XX
 FT Composition containing one or more DNA molecules encoding fragments of a
 FT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 FT tumor therapy.
 PT
 PT
 XX
 XX Claim 18; Fig 8; 56pp; English.
 PS
 XX The present sequence is a fusion protein consisting of human Mucin 1 (MUC
 CC -1) fragment fused to UBILacI sequence at the N-terminus. The UBILacI
 CC sequence consists of ubiquitin from MCF7 cell line and a portion of E.
 CC coli beta-galactosidase (lacI). MUC-1 is an antigenic protein
 CC overexpressed in tumour cells. The corresponding DNA sequence is cloned
 CC into a pMRS30 expression vector and used in pharmaceutical composition
 CC e.g. vaccine for inducing an antigen-specific anti-tumour immune
 CC response. Composition containing this DNA molecule is useful in anti-
 CC tumour therapy of patients affected with tumours characterised by high
 CC MUC-1 expression. (Updated on 12-SEP-2003 to standardise OS field)
 CC
 SQ Sequence 295 AA;
 Query Match 100.0%; Score 49; DB 3; Length 295;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHV 9
 DB 249 STAPPVHV 257
 RESULT 10
 ADA50571
 ID ADA50571 standard; protein; 307 AA.
 XX
 XX AC ADA50571;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Mucin 1 (MUC-1) splice variant #1, SEQ ID NO:26.
 XX
 KW Nucleic acid vaccine; DNA vaccine; tumour antigen; cytokine adjuvant;
 KW humoral response; cellular response; immune response; immunotherapy;
 KW cancer; cytostatic; vaccine; gene therapy; mucin 1; MUC-1.
 XX
 OS Unidentified.
 XX
 PN WO2003031569-A2.
 EN
 XX 17-APR-2003.
 PD
 XX


```

PF 18-SEP-2002; 2002WO-US029640.
XX
XX
XX 10-OCT-2001; 2001US-0328371P.
XX
XX (CENZ ) CENTOCOR INC.
XX
XX Snyder L, Scallon B, Knight DM, McCarthy SG, Goletz TJ;
PI Branigan PJ;
PI
XX
XX WPI; 2003-393437/37.
DR N-PSDB; ADA50572.
XX
XX New nucleic acid vaccine, useful for eliciting an immune response to a
PI cancer associated tumor protein in a mammal.
XX
XX Claim 1a; Page 38; 92pp; English.
XX
XX The invention relates to a nucleic acid vaccine comprising one or more
CC tumour antigen-encoding nucleic acids and one or more cytokine adjuvant-
CC encoding nucleic acids. The tumour antigen encoded by the vaccine is
CC mucin 1 (MUC-1), the kallikrein Kk2, or prostate specific antigen (PSA,
CC also known as KLK3), and the cytokine adjuvant encoded can be interleukin
CC -12 (IL-12), granulocyte macrophage-colony stimulating factor (GM-CSF),
CC or especially interleukin-18 (IL-18). The antigen-encoding nucleic acid
CC is preferably under the control of a promoter such as the cytomegalovirus
CC immediate early promoter, the dihydrofolate reductase promoter or the
CC early or late SV40 promoters. The invention also encompasses the method
CC of eliciting an immune response to a tumour antigen in a mammal using the
CC vaccine of the invention. Coexpression of the antigen and adjuvant
CC induces a humoral or cellular response to the tumour antigen, generating
CC an immune response useful for treatment or prophylaxis of cancers. The
CC present sequence represents a mucin 1 (MUC-1) polypeptide sequence which
CC is specifically claimed for use in the vaccine of the invention.
XX
XX
XX SQ Sequence 307 AA;
Query Match 100.0%; Score 49; DB 6; Length 307;
Best Local Similarity 100.0%; Pred. No. 2.7; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
QY 1 STAPPVHV 9
DB 170 STAPPVHV 178
|||||
|||||

RESULT 11
AAU84810
ID AAU84810 standard; protein; 312 AA.
XX
XX AAU84810;
AC
XX
XX 08-MAY-2002 (first entry)
DT
XX
XX Human MUC1R consensus sequence.
DE
XX
XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW viral infection; human immunodeficiency virus; melanoma;
KW bacterial infection; Salmonella; Legionella; parasitic infection;
KW Trypanosoma; Toxoplasma; Giardia.
XX
XX Homo sapiens.
OS
XX WO200190197-A1.
XX
XX 29-NOV-2001.
PD
XX
XX 25-MAY-2001; 2001WO-AU000622.
PF
XX
XX 26-MAY-2000; 2000AU-00007761.
PR
XX (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Thomson SA, Ramshaw IA;
PI

```

```

XX
DR WPI; 2002-147575/19.
XX
XX New synthetic polypeptides having several different segments of at least
PT one parent polypeptide linked together differently compared to the
PT linkage in the parent polypeptide, for inducing immune response against a
PT pathogen or cancer.
XX
XX Example 3; Fig 27; 364pp; English.
XX
XX The invention relates to a new synthetic polypeptide (I) comprising
CC several different segments of at least one parent polypeptide linked
CC together in a different relationship relative to their linkage in the
CC parent polypeptide to impede, abrogate or otherwise alter at least one
CC function associated with the parent polypeptide and for inducing an
CC immune response against a pathogen or cancer. Also included are a
CC synthetic polynucleotide encoding and a computer system for designing the
CC synthetic polynucleotides. The synthetic polypeptides and polynucleotides
CC are referred to as a Savine. The synthetic polypeptide is useful for a
CC modulating immune responses preferably directed against a pathogen or a
CC cancer. (e.g., cancers of the lung, breast, ovary, cervix, colon, head
CC and neck, pancreas, prostate, stomach, bladder, kidney, bone liver,
CC oesophagus, brain, testicle, uterus), as potentiating agents.
CC Compositions comprising the polypeptide may be used in the treatment or
CC prophylaxis against viral (such as infections caused by HIV (human
CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
CC Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is
CC a consensus sequence for a parent protein used to design a savine of the
CC invention
XX
XX SQ Sequence 312 AA;
Query Match 100.0%; Score 49; DB 5; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.8; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
QY 1 STAPPVHV 9
DB 7 STAPPVHV 15
|||||
|||||

RESULT 12
AAR96298
ID AAR96298 standard; peptide; 327 AA.
XX
XX AAR96298;
AC
XX
XX 26-JUL-1996 (first entry)
DT
XX
XX Glycoprotein 39 C terminal fragment.
DE
XX
XX Glycoprotein 39; gp39; lambda gt11 cDNA library; gastric cancer;
KW cell line KATO-III; tumour; immune abnormality; marker;
KW inflammatory disease.
XX
XX Homo sapiens.
OS
XX JF07051065-A.
XX
XX 28-FEB-1995.
PD
XX
XX 21-FEB-1992; 92JP-00035085.
PF
XX
XX 21-FEB-1992; 92JP-00035085.
PR
XX (NIKO-) NIPPON KOTAI KENKYUSHO KK.
XX (UYKA-) UNIV KAGOSHIMA.
XX
XX WPI; 1995-127356/17.
DR

```


CC protein overexpressed in tumour cells. The sequence was obtained from
 CC Br20 tumour cells. The corresponding DNA sequence is cloned into a pMRS30
 CC expression vector and used in pharmaceutical composition e.g. vaccine for
 CC inducing an antigen-specific anti-tumour immune response. Composition
 CC containing this DNA molecule is useful in anti-tumour therapy of patients
 CC affected with tumours characterised by high MUC-1 expression
 XX
 SQ Sequence 455 AA;
 Query Match 100.0%; Score 49; DB 3; Length 455;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHNV 9
 Db 150 STAPPVHNV 158
 Search completed: July 28, 2004, 18:09:30
 Job time : 53 secs

XX (TRGE) TRANSGENE SA.
 XX Chambon P, Kiemy MP, Lathe R, Hareuveni M;
 XX WPI; 1992-167097/20.
 DR N-PSDB; AAQ29276.
 XX
 XX Compsns. contg. polypeptide antigen recognised by antibody H23 - for
 PT treatment of mammary tumours, also for pox virus compsns. for use in
 PT vaccines.
 XX
 PS Claim 3; Page 15-18; 29pp; French.
 XX
 CC The tumour antigen recognised by antibody H23 is aberrantly expressed in
 CC epithelial cells from cancerous mammary tissue in about 90 per cent of
 CC breast cancer cases; in a normal individual expression is negligible. The
 CC antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S).
 CC Both forms show a high degree of polymorphism. A 20 amino acid subunit in
 CC ETA can be tandemly repeated up to 80 times. From one subunit to the
 CC next, 1 to 3 amino acids can differ. DNA coding for immunogenic fragments
 CC of ETA can be inserted into e.g. vaccinia viruses for treatment of
 CC mammary tumours. See also AAQ24678-Q24681, AAQ29277 and AAR23974-R23981.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 455 AA;
 Query Match 100.0%; Score 49; DB 2; Length 455;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHNV 9
 Db 150 STAPPVHNV 158
 RESULT 15
 AAY71024
 ID AAY71024 standard; protein; 455 AA.
 XX
 AC AAY71024;
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Human Mucin 1 (MUC-1) protein fragment #5.
 XX
 KW Human; Mucin 1; MUC-1; tumour; pMRS30 expression vector; anti-tumour;
 KW therapy; immune response; cytostatic; vaccine.
 XX
 OS Homo sapiens.
 XX
 XX WO200025827-A2.
 PN
 XX PD 11-MAY-2000.
 XX
 PF 18-OCT-1999; 99WO-EP007874.
 XX
 PR 30-OCT-1998; 98IT-MI002330.
 XX
 PA (MENA) MENARINI RICERCHE SPA.
 XX
 PI Parente D, Di Massimo AM, De Santis R;
 XX
 DR WPI; 2000-365410/31.
 DR N-PSDB; AAD00388.
 XX
 XX Composition containing one or more DNA molecules encoding fragments of a
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 PT tumor therapy.
 XX
 XX Claim 16; Fig 5; 56pp; English.
 PS
 XX The present sequence is a fragment of human Mucin 1 (MUC-1), an antigenic